

Thu Feb 20 14:41:30 2003

us-09-873-224-147.rge

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 18, 2003, 21:29:20 : Search time 1311 Seconds
(without alignments)
7658.633 Million cell updates/sec

Title: US-09-873-224-147
Sequence: 1 ATAGGACACTTCTTCAACCC.....AAATGACCCCGGCGCAGGA 345

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_un:*
28: em_vl:*
29: em_vl:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pin:*
35: em_hg_rod:*
36: em_hg_mam:*
37: em_hg_vrt:*
38: em_sy:*
39: em_hggo_hum:*
40: em_hggo_mus:*
41: em_hggo_other:*

pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	345	100.0	345	6	A40747	A40747 Sequence 14
2	345	100.0	345	6	AX031725	AX031725 Sequence
3	345	100.0	345	6	AX031995	AX031995 Sequence
4	345	100.0	345	6	AX032265	AX032265 Sequence
5	319.2	92.5	1584	14	HPCKR030A3	D49747 Hepatitis C
6	319.2	92.5	1584	14	HPCKR049E1	D49749 Hepatitis C
7	319.2	92.5	1584	14	HPCKR072A9	D49753 Hepatitis C
8	317.6	92.1	1584	14	HPCKR070A8	D49752 Hepatitis C
9	314.4	91.1	1584	14	HPCKR055A6	D49750 Hepatitis C
10	314.4	90.7	1584	14	HPCKR055A6	A50394 Sequence 49
11	312.8	89.6	309	6	AR127535	AR127535 Sequence
12	309	89.6	309	6	AR127535	AR127535 Sequence
13	309	89.6	309	6	AR127535	AR127535 Sequence
14	269.6	78.1	573	14	AF216792	AF216792 Hepatitis
15	269.6	78.1	573	14	AF216793	AF216794 Hepatitis
16	269.6	78.1	573	14	AF216794	AF216794 Hepatitis
17	269.6	78.1	573	14	HC033437	X76918 Hepatitis C
18	262.6	76.1	9390	14	HC033437	X76918 Hepatitis C
19	262.6	75.8	652	6	AS0404	A50404 Sequence 59
20	261.6	75.8	652	6	AR127540	AR127540 Sequence
21	261.6	75.8	652	6	HPCCOPRAB	L38349 Hepatitis C
22	261.6	75.8	652	6	HPCCOPRAB	A40765 Sequence 16
23	260.2	75.4	499	6	AX031743	AX031743 Sequence
24	260.2	75.4	499	6	AX032013	AX032013 Sequence
25	260.2	75.4	499	6	AX032013	AX032013 Sequence
26	260.2	75.4	499	6	AX032013	AX032013 Sequence
27	260.2	75.4	499	6	AX032013	AX032013 Sequence
28	260.2	75.4	499	6	AX032013	AX032013 Sequence
29	260.2	75.4	499	6	AX032013	AX032013 Sequence
30	259.4	75.2	573	14	HC010210	U10210 Hepatitis C
31	259.4	75.2	573	14	AF046866	AF046866 Hepatitis C
32	259.4	75.2	573	14	AF046866	AF046866 Hepatitis C
33	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
34	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
35	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
36	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
37	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
38	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
39	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
40	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
41	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
42	257.8	74.7	372	14	HC010235	U10235 Hepatitis C
43	256.8	74.3	573	14	HC010197	U10197 Hepatitis C
44	256.8	74.3	573	14	HC010197	U10197 Hepatitis C
45	256.8	74.3	573	14	HC010197	U10197 Hepatitis C

ALIGNMENTS

RESULT 1
LOCUS A40747 345 bp DNA
DEFINITION Sequence 147 from Patent WO9425601.
ACCESSION A40747
VERSION A40747.1 GI:2296744
KEYWORDS
SOURCE "unidentified."
ORGANISM "unidentified."
REFERENCE 1 (bases 1 to 345)
AUTHORS Maertens,G. and Stuyver,L.
TITLE NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
THERAPEUTIC AND DIAGNOSTIC AGENTS
JOURNAL Patent: WO 9425601-A 147 10-NOV-1994;

COMMENT

INNOGENETICS NV (BE)
 Other publication AU 2139100 941110
 Other publication CA 6722294 941121
 Other publication CN 1108030 950906
 Other publication FI 946066 941223
 Other publication NO 944967 941221
 Other publication JP 7508423 950921.

FEATURES

source
 1..345
 /organism="unidentified"
 /db_xref="taxon:32644"
 1..>345
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAA02529.1"
 /db_xref="GI:2296745"
 /translation="MSTLPKPKRKTNRNTPGHTLSSQAARSLVEFTCYHAGAPSW
 VCVQCARLPSPGRNLAVGANPSPGRAEPRAGPLSPGTLGPYGMRAAGGGSCPRRA
 LARRGAQMTPEAG"
 mat_peptide
 1..342
 /product="unnamed"

BASE COUNT 71 a 113 c 110 g 51 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 345; DB 6; Length 345;
 Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCACACTTCTTAACCAACCAAGAAAAACCAAGAAACCAACCAACCCCGGCACAGG 60
 Db 1 ATGAGCACACTTCTTAACCAACCAAGAAAAACCAAGAAACCAACCAACCCCGGCACAGG 60
 QY 61 ACCTTAAGTTCCAGAGCGCGGCGTCAAGTCGTTGGTGAAGTTACGTGCTACACGACAGG 60
 Db 61 ACCTTAAGTTCCAGAGCGCGGCGTCAAGTCGTTGGTGAAGTTACGTGCTACACGACAGG 60
 QY 121 GCCCCAGTTGGTGGTCTGTCAGTCGCAAGACTTCCGAGCGGTCCGCAACCTTGCAGTA 120
 Db 121 GCCCCAGTTGGTGGTCTGTCAGTCGCAAGACTTCCGAGCGGTCCGCAACCTTGCAGTA 120
 QY 181 GGGCCCAACCCATCCCAAGAGGCGCGGCGCAACGAGGAGGTCTCTGGGCTCAGCCGGGT 240
 Db 181 GGGCCCAACCCATCCCAAGAGGCGCGGCGCAACGAGGAGGTCTCTGGGCTCAGCCGGGT 240
 QY 241 ACCCTTGGCCCTATATAGGAATGAGGCTGCGGGTGGGAGGGTCTGCTCCCGGC 300
 Db 241 ACCCTTGGCCCTATATAGGAATGAGGCTGCGGGTGGGAGGGTCTGCTCCCGGC 300
 QY 301 GGGGCTCTGCGCGTCTGTTGGGCGCCAAATGATGACCCCGGCGCAGGA 345
 Db 301 GGGGCTCTGCGCGTCTGTTGGGCGCCAAATGATGACCCCGGCGCAGGA 345

RESULT 2

LOCUS AX031725 345 bp DNA linear PAT 20-SEP-2000
 DEFINITION Sequence 147 from Patent EP1004670.
 ACCESSION AX031725
 VERSION AX031725.1 GI:10278924
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 345)
 AUTHORS Maertens G. and Stuyver L.
 TITLE Sequences of hepatitis C virus genotypes and their use as
 JOURNAL therapeutic and diagnostic agents
 INNOGENETICS NV (BE)
 Patent: EP 1004670-A 147 31-MAY-2000;
 LOCATION/Qualifiers
 1..345
 /organism="unidentified"
 /db_xref="taxon:32644"

FEATURES
 source
 1..345
 /organism="unidentified"
 /db_xref="taxon:32644"

CDS

1..>345
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC09707.1"
 /db_xref="GI:10278925"
 /translation="MSTLPKPKRKTNRNTPGHTLSSQAARSLVEFTCYHAGAPSW
 VCVQCARLPSPGRNLAVGANPSPGRAEPRAGPLSPGTLGPYGMRAAGGGSCPRRA
 LARRGAQMTPEAG"
 mat_peptide
 1..342
 /product="unnamed"

BASE COUNT 71 a 113 c 110 g 51 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Score 345; DB 6; Length 345;
 Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCACACTTCTTAACCAACCAAGAAAAACCAAGAAACCAACCAACCCCGGCACAGG 60
 Db 1 ATGAGCACACTTCTTAACCAACCAAGAAAAACCAAGAAACCAACCAACCCCGGCACAGG 60
 QY 61 ACCTTAAGTTCCAGAGCGCGGCGTCAAGTCGTTGGTGAAGTTACGTGCTACACGACAGG 60
 Db 61 ACCTTAAGTTCCAGAGCGCGGCGTCAAGTCGTTGGTGAAGTTACGTGCTACACGACAGG 60
 QY 121 GCCCCAGTTGGTGGTCTGTCAGTCGCAAGACTTCCGAGCGGTCCGCAACCTTGCAGTA 120
 Db 121 GCCCCAGTTGGTGGTCTGTCAGTCGCAAGACTTCCGAGCGGTCCGCAACCTTGCAGTA 120
 QY 181 GGGCCCAACCCATCCCAAGAGGCGCGGCGCAACGAGGAGGTCTCTGGGCTCAGCCGGGT 240
 Db 181 GGGCCCAACCCATCCCAAGAGGCGCGGCGCAACGAGGAGGTCTCTGGGCTCAGCCGGGT 240
 QY 241 ACCCTTGGCCCTATATAGGAATGAGGCTGCGGGTGGGAGGGTCTGCTCCCGGC 300
 Db 241 ACCCTTGGCCCTATATAGGAATGAGGCTGCGGGTGGGAGGGTCTGCTCCCGGC 300
 QY 301 GGGGCTCTGCGCGTCTGTTGGGCGCCAAATGATGACCCCGGCGCAGGA 345
 Db 301 GGGGCTCTGCGCGTCTGTTGGGCGCCAAATGATGACCCCGGCGCAGGA 345

RESULT 3

LOCUS AX031995 345 bp DNA linear PAT 20-SEP-2000
 DEFINITION Sequence 147 from Patent EP0984068.
 ACCESSION AX031995
 VERSION AX031995.1 GI:10279107
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 345)
 AUTHORS Maertens G. and Stuyver L.
 TITLE Sequences of hepatitis C virus genotypes and their use as
 JOURNAL therapeutic and diagnostic agents
 INNOGENETICS NV (BE)
 Patent: EP 0984068-A 147 08-MAR-2000;
 LOCATION/Qualifiers
 1..345
 /organism="unidentified"
 /db_xref="taxon:32644"
 1..>345
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC09783.1"
 /db_xref="GI:10279108"
 /translation="MSTLPKPKRKTNRNTPGHTLSSQAARSLVEFTCYHAGAPSW
 VCVQCARLPSPGRNLAVGANPSPGRAEPRAGPLSPGTLGPYGMRAAGGGSCPRRA
 LARRGAQMTPEAG"
 mat_peptide
 1..342
 /product="unnamed"

BASE COUNT 71 a 113 c 110 g 51 t

ORIGIN

DB 6; Length 345;

Query Match	100.0%;	Score	60;
Best Local Similarity	100.0%;	Pred. No.	7,1e-70;
Matches	345;	Mismatches	0;
	Conservative	Indels	0;
		Gaps	0;

[illegible]

RESULT 4				
AX032265	345 bp	DNA	linear	PAT 20-SEP-2000

LOCUS	AX034263	Sequence 147 from Patent EP0984067.
DEFINITION	AX034265	
ACCESSION	AX034265.1	GI:10279290
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
		unidentified.
		unidentified
		unclassified.
REFERENCE		1 (bases 1 to 345)
AUTHORS		Maertens,O. and Stuyver,L.
TITLE		Sequences of hepatitis C virus genotypes and their use as
JOURNAL		therapeutics and diagnostic agents
		Patent: EP 0984067-A 147 08-MAR-2000;
		INNOGENETICS NV (BE)

Location/Qualities	FEATURES
1. .345	source
organism="unidentified"	

```

/ab_anc
1. :>345
"unnamed protein product"

```

```

/note= "start=1
/codon_start=1
/protein_id="CAC09860.1"
/db_xref="GI:10279291"
/translation="MSTLPKPKRKTNTNPGHRTLSOAAVRSLVETTCYHAGAPSM
VCVQCARLPSGRILAVGANPSPGAEPRAGGLSPGLTIPYMGMAAGGGGSCPRRA
LARGAQMTPGAG"

```

```

BASE COUNT      1.1 3
ORIGIN
nb 6: length 345;

```

Query Match	100.0%	Score 345	Id	0
Best Local Similarity	100.0%	Prod. No. 7	Indels	0
Matches 345		Conservative	Mismatches 0	Gaps 0

QY 61 ACGTTAAGTTCACGAGCGGGCTCAGATCGTTGGTAGAGTTACGTGCATTCCACGCAAGG 120

Db	61	ACCTTAGTTCACAGCGCGCGTCAGTCTCTGTGTGATTTACGTCTACACGACAGGG	120
Qy	121	GGCCCCAGTTGGGTGTGGTGTGACAGTCCGAGGCGGTGCGACACCTCCACATA	180
Db	121	GGCCCCAGTTGGGTGTGGTGTGACAGTCCGAGGCGGTGCGACACCTCCACATA	240
Qy	181	GGCGCCACCCATATCCCAAGGCGGCCCAACGAGGCGAGTCTGTGGGTCTAGCCGGGT	240
Db	181	GGCGCCACCCATATCCCAAGGCGGCCCAACGAGGCGAGTCTGTGGGTCTAGCCGGGT	240
Qy	241	GGCGCCACCCATATCCCAAGGCGGCCCAACGAGGCGAGTCTGTGGGTCTAGCCGGGT	300
Db	241	GGCGCCACCCATATCCCAAGGCGGCCCAACGAGGCGAGTCTGTGGGTCTAGCCGGGT	300
Qy	301	ACCCCTTGCCCTTATATGAGATGAGGGTGTGGGTGTGGAGGTGCCTCTGTCCCGC	360
Db	301	ACCCCTTGCCCTTATATGAGATGAGGGTGTGGGTGTGGAGGTGCCTCTGTCCCGC	360
Qy	361	GGCGCTCTGCGCCGTCGTGTGGGCCCAATATACCCCGCGGCAGGA	420
Db	361	GGCGCTCTGCGCCGTCGTGTGGGCCCAATATACCCCGCGGCAGGA	420

RESULT 5

linear

VRL 10-FEB-1999

HEPCJK030A03		1584 bp	nuc
TAXONOMY		Hepatitis C virus isolate JK030 gene for core, env, and part of NSI	
DEFINITION		E2/NSI, partial cds.	
ACCESSION	D49747	GI:1197102	
VERSION	D49747.1		
KEYWORDS		core, env, and part of E2/NSI.	
SOURCE		Hepatitis C virus (isolate:JK030) CDNA to genomic RNA.	
ORGANISM		Hepatitis C virus	
		Hepatitis C virus	
		Hepatitis C virus	
		ssRNA positive-strand viruses; no DNA stage; Flaviviridae;	

REFERENCE	AUTHORS	TITLE	JOURNAL	DATE
1	Okamoto, H., Iizuka, H., Kishimoto, J., Tsuda, F., Tokita, T., Miyakawa, Y., and Mayumi, M.	Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups	J. Gen. Virol.	77 (Pt 2), 293-301 (1996)
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				

MANUSCRIPT	
REFERENCE	2 (bases 1 to 1867)
AUTHORS	Okamoto, H.
JOURNAL	unpublished
REFERENCE	3 (bases 1 to 1584)
AUTHORS	Okamoto, H.
TITLE	Direct Submission
JOURNAL	Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School, Immunology Division; Mhamkakawachi-machi, Kawachi-gun, Tochigi 329-004 Japan (E-mail: hokamoto@jichi.ac.jp), 329-0285-44-2111(ex.3334), Fax:0285-44-1557)

FEATURES	Location/Qualities
source	1. .1584
	/organism="Hepatitis C virus"

```

<1. .000
5'UTR
/citation=[2]
/evidence=not_experimental
340. >1584
cns
01

```

[illegible]

BASE COUNT 308 a 459 c 453 g 364 t
 ORIGIN

Query Match

Best Local Similarity 92.5%; Score 319.2; DB 14; Length 1584;
 Matches 335; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

1 ATGAGCACACTTCTTAACCAAGAAAGAAACCAAAAGCAACCAAC-CCGGCCACAG 59
 340 ATGAGCACACTTCTTAACCAAGAAAGAAACCAAAAGCAACCAAC-CCGGCCACAG 59
 60 GACGTTAAGTCCAGAGCGCGGTCATGTTGTTGAGTTTACGCTTACACGACG 119
 400 GACGTTAAGTCCAGAGCGCGGTCATGTTGTTGAGTTTACGCTTACACGACG 119
 120 GACGTTAAGTCCAGAGCGCGGTCATGTTGTTGAGTTTACGCTTACACGACG 459
 460 GACGTTAAGTCCAGAGCGCGGTCATGTTGTTGAGTTTACGCTTACACGACG 459
 180 AGGCGCCAAACCCATCCCAAGAGCGCGGTCATGTTGTTGAGTTTACGCTTACACGACG 519
 520 AGGCGCCAAACCCATCCCAAGAGCGCGGTCATGTTGTTGAGTTTACGCTTACACGACG 239
 240 TACCTTGGCCCTTATGGAATGAGAGCGTGGGTCGAGAGGTCCTGCTCAGCCCGG 579
 580 TACCTTGGCCCTTATGGAATGAGAGCGTGGGTCGAGAGGTCCTGCTCAGCCCGG 299
 300 CGCGGCTCTCGCGGTCGAGAGGTCGAGAGGTCCTGCTCAGCCCGG 639
 640 CGCGGCTCTCGCGGTCGAGAGGTCGAGAGGTCCTGCTCAGCCCGG 683

RESULT 6

HPcJK049A5

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

Journal

Medline

Reference

Authors

Title

AUTHORS	TITLE
Tokita, H., Okamoto, H., and Mayumi, M.	Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and

eleventh (11a) genetic groups
Gen. Virol. 77 (Pt 2), 293-301 (1996)

JOURNAL
MEDLINE
REFERENCE
96226020
2 (bases 1 to 1584)

AUTHORS	Okamoto, H.
JOURNAL	unpublished
REFERENCE	3 (bases 1 to 1584

AUTHORS Okamoto, H.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi

329-04, Japan (E-mail: hokamoc@jcom.ne.jp)
Tel: 0285-44-2111 (ex. 3334), Fax: 0285-44-
1.1564
Location/Qualifiers
FEATURES
Source
"Hepatitis C virus"
organism="Hepatitis C virus"

```
5'UTR <1. .339  
/citation=[2]  
/evidence=not_experimental  
CDS 340. .>1584  
/citation=[2]
```

```

/codon_start=1
/evidence=not-experimental
/product="core, env, and part of Ez/NS1"
/protein_id="BA00586.1"
/db_xref="GI:1197163"
/db_xref="X01783"
/translation="WSTLPKPKORTKRTNRRRPODKRPFGRQIIVGVYVLPKRPGL
GVAHKTSEKSRQSPRPYPRAPRREGESMAQPCPEMPILATNGCCGMAWMLISG
SRSPGNDPNRRRNRNLICKYDITLTGCFALMKYPIVAGAPGVARALAGVALLD
GTFATGNGLPCCSEFSLPLALLSCLLTPTAGLEYRNVSGLITVNDCSNSIIEADY
IILHDCGYPCVSCGNTSRCTPVSPYPAVASRGVATSLRTHYDMWGAATLSEAVV
VSDLCGALFLVCGSGMRHRQHTVCECNSITPIGHITGRHMMNMNMNSPVTAVV
SOVLRIPOTFLDVAGAHWGMAGVAYYSQGMWAVFLVLCFSGVDASTIISGSGS
ARSWGITLSFSGSNONLQLYN"

```

BASE COUNT	ORIGIN
97.18;	Score 314.4; DB 14; Length 1584

Query Match	92.58;	Pred. No. 6,7e-63;	Indels	1;	Gaps	1
Best Local Similarity	96.58;					
Matches 332; Conservative	0;	Mismatches	11;			

1 ATGAGCACACTTCTTAACCACAAAGAAAAACCCAAAAGAAACACCAAC -CCGGCCACAG 399

Dd 340 ATGAGCACACATTCCTAAACCACCAAGAGAAAACCCATTGGTGGTTA
60 CACCGTAAGTCCCAGGCGGCGGTCCAGATCGTTGGTAGTTACGTACCACGAGC

[illegible]

QY 120 GGGCCCCAGTTGGGTGTGCGTGCAGTGGCGAAGACTTCGCCGGCGGTGCGACACCCCGCCTTG
TTTTTT TTTTTTTTCTCCGATGCAAGAAGACTTCCGGCGGTGCGAACCTCGCAGT 519

Dd 460 GGGCCAGATTGCGTGTCTCCTACCAATGCAGTTTCAGTG
 180 AGCGGCCAACCATCCCGCAGGCGCGCGAACCAGAGGCGACTCCTGGGCTCAGCCCCGCG 239

Ov 720

Db 520 AGGCGCAACCCATCCCCAGGCGCGCCGAACCTGAGG6CCGGTCTCGSACTAGCCCCG 299

[illegible]

Db	580	TACCC109cccccccccccccccc	343
QY	300	CGCGGCTCTGCCCCGTCGTGGGCCCAATGACCCCCGGCGAG	343

640 CCGGGCTCTCGCCCATCGTGGGGCCCAATGACCCCGGGGTAG 683

RESULT 11

		RNA	linear	VRL 13-JUL-1987
	325 bp			
HC3NL96	HC3NL96	VIRUS RNA genotype 3	NL96 core.	
LOCUS				
HLA-DQA1				

DEFINITION	hepatitis C
ACCESSION	X78863
VERSION	X78863.1 GI:474329

KEYWORDS core region, Hepatitis C virus, Hepatitis C virus, Hepatitis C virus, positive-strand viruses, no DNA stage: Flaviviridae

Viruses: ssRNA positive
Hepadnavirus
1 (bases 1 to 325)

REFERENCE	AUTHORS	TITLE
van Doorn, L.J.		
Direct Submission		
Submitted (15-APR-1994)	L.J. van Doorn	Diagnostic Centre SSDZ, G. van Doorn, PO Box 5010, 2600

Journal of Molecular Biology, R. de Gierweg 17, 2615
Dept. of Molecular Biology, R. de Gierweg 17, 2615
Delft, NETHERLANDS (325)

REFERENCE
1. van Doorn, L. J., Klever, G. E., Stuyver, L., Maatleens, R., Heijthink, R. A. and Quint, W. G. Schalm, S. W. Use of hepatitis C virus genotypes 1 to 5 reveals

TITLE Sequence analysis of hepatitis B virus multiple novel subtypes in the Benelux countries
J. Gen. Virol. 76 (Pt 7), 1871-1876 (1995)

location/Qualifiers

```

FEATURES
source
1. . 325
/organism="Hepatitis C virus"
/specific host="Homo sapiens"

```

```

/species="taxon:11103"
/db_xref="Human plasma"
/cell_line="Human plasma"

```

```
CDS
1..>323
/codon_start=1
/core_region
/product="core region"
"naastaf12 1"
```

```
/protein_id="CAA0312.2  
/db_xref="GI:474330"  
/db_xref="SPTREMBL:Q68494"  
/db_xref="XP_002842507.1:SPTRKNTNXXRPDVKFPPGSGQLVGGYVLPRPG
```

```

/translation="MSIL;PUN;I;N;A;V;E;R;S;E;
GVRAVAKT;E;S;Q;P;R;S;R;Q;P;I;P;R;A;R;T;E;G;R;S;W;A;Q;P;G;Y;P;W;P;L;Y;N;E;G;G;C;W;A;G;W;L;L;S;
SRPSWG"
to 1 others

```

BASE COUNT	BASE COUNT	BASE COUNT	BASE COUNT
65 a	105 C	104 g	50 L

Query Match	90.78%	Score 313;	DB 14;	Length 323;
Best local Similarity	99.78%	Pred. No. 1.9e-62;		
		Mismatches 0;	Indels 1;	Gaps

Matches	324;	Conservative	0;	Mismatch	59
1	ATACGACACTTCTTAACCA	CAAGAAAAACCAAGAACACCAACC	-CCGGCCACAG	59	

Db

1 ATGAGCACACTTCTTAACCAACAAGAAAACCACCAACCSCCGGCACAG 60
119

[illegible]

Db
61 GACGTTAAGTTCACAGGCGCCGCTGGATTCGAGCGGTGCACAATCCGCACT 179
120 GGCCCCAGTTCGGGTGTCGTCGATGCCAGACTTCCGAGCAGGTGCACAATCCGCACT 180

121	GGCCCCAGTTGGGTGTGGGTGACAGTGGCAAGACTTTCGAGCGGTCCGAACTTCGAGT	186
Db		

[illegible]

D_b 181 AGAGGCGCAGCCGGCTGCCTGTCTCCCG 299
D_c 240 TACCCCTGGCCCCATATGGGAATGAGGGCTGGCGGTGGCAGCGTGGCTCTCTCCCG 300

Db 241 TACCCTGGCCCCATATGGAATGAGGGCTGCGGGTGGGCAGGGTGCGTCCTCCTGCCCC

QY	300	CGCGGCTCTGCCCCGTCGCGGACC	227
301	CGCGGCTCTGCCCCGTCGCGGACC <td>325</td> <td></td>	325	

RESULT 12
HPCJK055A6

100

LOCUS HPCJK055A6 1584 bp RNA linear VR1 10-FEB-1999
 DEFINITION Hepatitis C virus isolate JK055 gene for core, env, and part of
 E2/NS1, Partial cds.
 ACCESSION D49750
 VERSION D49750.1 GI:1197114
 KEYWORDS core, env, and part of E2/NS1.
 SOURCE Hepatitis C virus (isolate:JK055) CDNA to genomic RNA.
 ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepelivirinae.

REFERENCE 1 (sites)
 AUTHORS Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F.,
 Lesmana,U.A., Miyakawa,Y. and Mayumi,M.
 TITLE Hepatitis C virus variants from Jakarta, Indonesia classifiable
 into novel genotypes in the second (2e and 2f), tenth (10a) and
 eleventh (11a) genetic groups
 J. Gen. Virol. 77 (Pt 2), 293-301 (1996)

JOURNAL MEDLINE
 REFERENCE 2 (bases 1 to 1584)
 AUTHORS Okamoto,H.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1584)
 AUTHORS Okamoto,H.
 TITLE
 JOURNAL Direct Submission
 Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
 Immunology Division, Minamikawachi-machi, Kawachi-gun, Tochigi
 329-04, Japan (E-mail:hokamoto@jichi.ac.jp,
 Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)
 Location/Qualifiers
 1..1584
 /organism="Hepatitis C virus"
 /isolate="JK055"
 /db_xref="taxon:11103"
 <1..339
 /citation=[2]
 /evidence=not_experimental
 340..>1584
 /citation=[2]
 /codon_start=1
 /evidence=not_experimental
 /product="core, env, and part of E2/NS1"
 /protein_id="BA08584.1"
 /db_xref="GI:1197115"
 /translation="MSTLPKPPORTKRNTRRNPQVKKFPGSGQIVGVYVLPKRPKL
 GYRAVKTSESRSPRRPPIPRARPRERMAQDEYWPVLGNGCGMAGLLSPG
 SPSWGPNDPRRSRLNGKYIDLTGCGFADLNGTIPVGAIVGVAALAGVLEAD
 IILHPCGVCVRAGNTSKCTPISTFVAVSRPAATSLRTHDMVGAATLCSALY
 VGDLCALFTLVGGESFMRHOMHTVDDCSLYPGLTHGMDMMNMSPAYTNV
 SOVLRPOTLIDLVIGAHNGVMAGVAAYVSMGNMAKVFVLVCLPSGVDAITRTIAGSA
 AASTRGASLFTTGARONQLVNA"

FEATURES
 source
 1..1584
 /organism="Hepatitis C virus"
 /isolate="JK055"
 /db_xref="taxon:11103"
 <1..339
 /citation=[2]
 /evidence=not_experimental
 340..>1584
 /citation=[2]
 /codon_start=1
 /evidence=not_experimental
 /product="core, env, and part of E2/NS1"
 /protein_id="BA08584.1"
 /db_xref="GI:1197115"
 /translation="MSTLPKPPORTKRNTRRNPQVKKFPGSGQIVGVYVLPKRPKL
 GYRAVKTSESRSPRRPPIPRARPRERMAQDEYWPVLGNGCGMAGLLSPG
 SPSWGPNDPRRSRLNGKYIDLTGCGFADLNGTIPVGAIVGVAALAGVLEAD
 IILHPCGVCVRAGNTSKCTPISTFVAVSRPAATSLRTHDMVGAATLCSALY
 VGDLCALFTLVGGESFMRHOMHTVDDCSLYPGLTHGMDMMNMSPAYTNV
 SOVLRPOTLIDLVIGAHNGVMAGVAAYVSMGNMAKVFVLVCLPSGVDAITRTIAGSA
 AASTRGASLFTTGARONQLVNA"

BASE COUNT 311 a 467 c 453 g 351 t
 ORIGIN

Query Match 90.7%; Score 312.8; DB 14; Length 1584;
 Best Local Similarity 96.2%; Pred. No. 1.6e-62;
 Matches 331; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1 ATGAGCACACTTCTTAACACAAAGAAAACCAAAACCAACAC-CCGGCCACAG 59
 Db 340 ATGAGCACACTTCTTAACACAAAGAAAACCAAAACCAACAC-CCGGCCACAG 59
 QY 60 GAGCTTAAGTTCCAGCGCGCGGTGAGATCGTTGGTGAAGTTTACGCTACACGACAG 399
 Db 400 GAGCTTAAGTTCCAGCGCGCGGTGAGATCGTTGGTGAAGTTTACGCTACACGACAG 399
 QY 120 GGGCCCAAGTTGGGTGTCAGTGCAGTTCGAGGCGTTCGACACGACAG 459
 Db 460 GGGCCCAAGTTGGGTGTCAGTGCAGTTCGAGGCGTTCGACACGACAG 459
 QY 180 AAGCCCAACCAACGAGGCGCGCAACGAGGCGAGGCTCGGGCTAGCCGCGG 239
 Db 520 AAGCCCAACCAACGAGGCGCGCAACGAGGCGAGGCTCGGGCTAGCCGCGG 239

QY 240 TACCTTGCCCTTATATGGAATGAGGCTGGGGTGGGACAGGTGCTCTGTCGCC 299
 Db 580 TACCTTGCCCTTATATGGAATGAGGCTGGGGTGGGACAGGTGCTCTGTCGCC 299
 QY 300 CCGGCTCTCGCCCGCTGTCGGGGCCCAATGATACCCCGGCGAG 343
 Db 640 CCGGCTCTCGCCCGCTGTCGGGGCCCAATGATACCCCGGCGAG 343

RESULT 13
 LOCUS A50394
 DEFINITION Sequence 49 from Patent W09613590.
 ACCESSION A50394
 VERSION A50394.1 GI:2303406
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 309)
 AUTHORS Maertens,G. and Stuyver,L.
 TITLE NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
 JOURNAL PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS
 INNOVATION NV (BE)
 Patent: WO 9613590-A 49 09-MAY-1996;
 other publication AU 3844095 960523.
 COMMENT
 FEATURES
 source
 1..309
 Location/Qualifiers
 1..309
 /organism="unidentified"
 /db_xref="taxon:32644"

BASE COUNT 65 a 99 c 97 g 48 t
 ORIGIN
 Query Match 89.6%; Score 309; DB 6; Length 309;
 Best Local Similarity 100.0%; Pred. No. 1.6e-61;
 Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCACACTTCTTAACACAAAGAAAACCAAAACCAACAC-CCGGCCACAG 60
 Db 1 ATGAGCACACTTCTTAACACAAAGAAAACCAAAACCAACAC-CCGGCCACAG 60
 QY 61 AGCTTAAGTTCCAGCGCGCGGTGAGATCGTTGGTGAAGTTTACGCTACACGACAGG 120
 Db 61 AGCTTAAGTTCCAGCGCGCGGTGAGATCGTTGGTGAAGTTTACGCTACACGACAGG 120
 QY 121 GGGCCCAAGTTGGGTGTCGTCGAGTCGCAAGACTTCGAGGCGGTGCGAACCTGACAGTA 180
 Db 121 GGGCCCAAGTTGGGTGTCGTCGAGTCGCAAGACTTCGAGGCGGTGCGAACCTGACAGTA 180
 QY 181 GGGCCCAACCATCCAGGCGCGCGCGCAACCGAGGCGAGTCTGCGTACCGCGGAT 240
 Db 181 GGGCCCAACCATCCAGGCGCGCGCGCAACCGAGGCGAGTCTGCGTACCGCGGAT 240
 QY 241 ACCCTTGCGCCCTATATGGAATGAGGCTCGGGTGGGCGAGGCTGCTCTGTCGCCG 300
 Db 241 ACCCTTGCGCCCTATATGGAATGAGGCTCGGGTGGGCGAGGCTGCTCTGTCGCCG 300
 QY 301 GCGGCTCTC 309
 Db 301 GCGGCTCTC 309

RESULT 14
 LOCUS ARI27535
 DEFINITION Sequence 49 from patent US 6180768.
 ACCESSION ARI27535
 VERSION ARI27535.1 GI:14114130
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.

REFERENCE 1 (bases 1 to 309)
 AUTHORS Maertens, G., and Stuyver, L.
 TITLE Sequences of hepatitis C virus genotypes and their use as
 JOURNAL Propylactic, therapeutic and diagnostic agents
 FEATURES Patent: US 6180768-A 49 30-JAN-2001;
 source Location/Qualifiers
 1.309
 /organism="unknown"
 BASE COUNT 65 a 99 c 97 g 48 t
 ORIGIN

Query Match 89.6%; Score 309; DB 6; Length 309;
 Best Local Similarity 100.0%; Pred. No. 1.0e-61;
 Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCACACTTCTTAACCAAGAAAAACCAAAACCAACACCCGCGCACAGG 60
 Db 1 ATGAGCACACTTCTTAACCAAGAAAAACCAAAACCAACACCCGCGCACAGG 60
 QY 61 ACCTTAAGTTCCAGCGCGCGGTCAAGATCGTTGGTGAAGTTACGTGCTACACAGCAGG 120
 Db 61 ACCTTAAGTTCCAGCGCGCGGTCAAGATCGTTGGTGAAGTTACGTGCTACACAGCAGG 120
 QY 121 GCCCCAGTTGGGTGGCGTGCAGTGGCAGAGCTTCGAGCGGTCCGTAACCTGCGAGTA 180
 Db 121 GCCCCAGTTGGGTGGCGTGCAGTGGCAGAGCTTCGAGCGGTCCGTAACCTGCGAGTA 180
 QY 181 GGGCCCAACCCATCCCAAGGCGCGCCGACACGAGGCGAGTCCGAGCGGTCCGAGT 240
 Db 181 GGGCCCAACCCATCCCAAGGCGCGCCGACACGAGGCGAGTCCGAGCGGTCCGAGT 240
 QY 241 ACCCTTGGCCCTATATAGGAATGAGGCTCGGGGTGGGCGAGGCTGCTCTGCCCCG 300
 Db 241 ACCCTTGGCCCTATATAGGAATGAGGCTCGGGGTGGGCGAGGCTGCTCTGCCCCG 300
 QY 301 GCGGCTCTC 309
 Db 301 GCGGCTCTC 309

RESULT 15
 AF216792 573 bp RNA linear VRL 26-JUN-2000
 LOCUS Hepatitis C virus isolate SOM1 core structural protein gene,
 DEFINITION partial cds.
 ACCESSION AF216792
 VERSION AF216792.1 GI:7274362
 KEYWORDS
 SOURCE Hepatitis C virus.
 ORGANISM Hepatitis C virus.
 Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 REFERENCE 1 (bases 1 to 573)
 AUTHORS Abid, K., Quadri, R., Veuthey, A.L., Hadenque, A. and Negro, F.
 TITLE A novel hepatitis C virus (HCV) subtype from Somalia and its
 JOURNAL classification into HCV clade 3
 MEDLINE J. Gen. Virol. 81 (Pt 6), 1485-1493 (2000)
 PUBMED 20273994
 10811932
 REFERENCE 2 (bases 1 to 573)
 AUTHORS Abid, K., Quadri, R. and Negro, F.
 TITLE Direct Submission
 JOURNAL Submitted (16-DEC-1999) Gastroenterology and Hepatology, University
 Hospital, 24 Rue Micheli-du-Crest, Geneva 1211, Switzerland
 FEATURES
 source
 1.573
 /organism="Hepatitis C virus"
 /isolate="SOM1"
 /db_xref="taxon:11103"
 /country="Somalia"
 /note="Clade 3
 subtype: 3h"
 1..>573

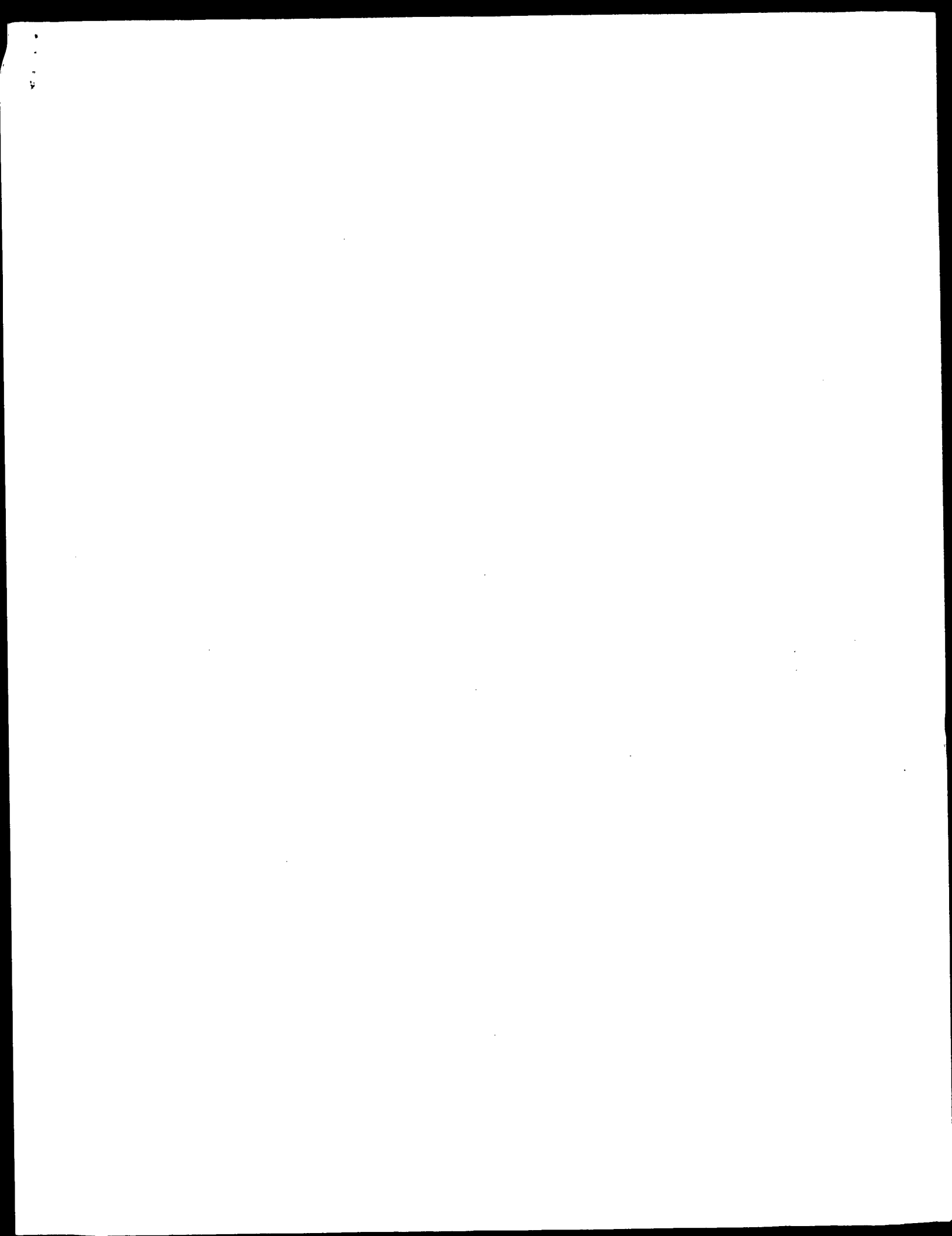
CDS

/codon_start=1
 /product="core structural protein"
 /protein_id="AA044739.1"
 /db_xref="GI:7274362"
 /translation="MSTLPKPKRKTKRNTIRPQNVKFPGGGQIVGVYVLPKPTL
 GYRAARKTSERSOPRGKRPVPAKRNKGRVMAQPGVPMPLYNKGGAGWLLSPRG
 SPPHNGPNDPRRSNLSKIIDLTLCGFADLMGIPLVGAPVGVARALAHGVRAVED
 GINTATGWLPCGSFIFLALTLSCITVPASG"
 BASE COUNT 110 a 179 c 172 g 112 t
 ORIGIN

Query Match 78.1%; Score 269.6; DB 14; Length 573;
 Best Local Similarity 88.4%; Pred. No. 2e-52;
 Matches 304; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

QY 1 ATGAGCACACTTCTTAACCAAGAAAAACCAAAAGAACCAACACC-CGCGCACAG 59
 Db 1 ATGAGCACACTTCTTAACCAAGAAAAACCAAAAGAACCAACACC-CGCGCACAG 60
 QY 60 GACGTTAAGTTCCAGCGCGCGGTGAGATCGTTGGTGAAGTTTACGTGCTACACAGCAGG 119
 Db 61 AACGTAAGTTCCAGCGCGCGGTGAGATCGTTGGTGAAGTTTACGTGCTACACAGCAGG 120
 QY 120 GCCCCAGTTGGGTGGCGTGCAGTGGCAGAGCTTCGAGCGGTCCGTAACCTGCGAGT 179
 Db 121 GCCCCAGTTGGGTGGCGTGCAGTGGCAGAGCTTCGAGCGGTCCGTAACCTGCGAGT 180
 QY 180 AGCGCCCAACCCATCCCAAGGCGCGCCGAAACGAGGCGAGTCTCGGCTCAAGCCGAG 239
 Db 181 AGCGCCCAACCCATCCCAAGGCGCGCCGAAACGAGGCGAGTCTCGGCTCAAGCCGAG 240
 QY 240 TACCTTGGCCCTATATAGGAATGAGGCTCGCGGTGGGCGAGGCTGCTCTGCCCCG 299
 Db 241 TACCTTGGCCCTATATAGGAATGAGGCTCGCGGTGGGCGAGGCTGCTCTGCCCCG 300
 QY 300 GCGGCTCTCGCCCGCTGCTGGGCGCCCAATGACCCCGGCGCAG 343
 Db 301 GCGGCTCTCGACCACTGGGCGCCCAATGACCCCGGCGCAG 344

Search completed: February 19, 2003, 00:50:09
 Job time : 1317 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2003, 21:11:31 ; Search time 241 Seconds
(without alignments)
3223.815 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 345
Sequence: 1 ATGACGACACCTTCTTAAC...AATGACCCCGCGCAGCA 345

Scoring table: IDENTITY_MDC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq-101002:*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	345	100.0	345	15	AA078089
2	298	86.4	310	17	AA078091
3	261.6	75.8	652	17	AA078096
4	260.2	75.4	499	15	AA078098
5	259.8	75.3	552	20	AA078099
6	259.4	75.2	573	17	AA078100
7	257.8	74.7	573	17	AA078101
8	257.8	74.7	803	15	AA078102
9	257.8	74.7	803	15	AA078103

10	257.4	74.6	552	20	AA078104	HCV J7 C/E domain
11	257	74.5	552	12	AA011075	Fragment of hepati
12	256.2	74.3	552	16	AA079749	Hepatitis C virus
13	256.2	74.3	552	16	AA079756	Hepatitis C virus
14	256.2	74.3	552	20	AA026732	Consensus sequence
15	256.2	74.3	552	20	AA026727	Consensus sequence
16	256.2	74.3	552	20	AA000400	Hepatitis C virus
17	256.2	74.3	570	13	AA021748	HCV clone HCV JH C
18	256.2	74.3	573	17	AA016642	Hepatitis C virus
19	256.2	74.3	573	17	AA016644	Hepatitis C virus
20	256.2	74.3	573	17	AA016645	Hepatitis C virus
21	256.2	74.3	1765	16	AA079141	Hepatitis C virus
22	254.6	73.8	499	15	AA078097	HCV core-envelope
23	254.6	73.8	531	13	AA032438	Hepatitis C virus
24	254.6	73.8	552	16	AA079755	Hepatitis C virus
25	254.6	73.8	573	17	AA016614	Hepatitis C virus
26	254.6	73.8	573	17	AA016621	Hepatitis C virus
27	254.6	73.8	573	17	AA016626	Hepatitis C virus
28	253.6	73.5	498	15	AA078112	Hepatitis C virus
29	253.6	73.5	573	17	AA016646	Hepatitis C virus
30	253.6	73.5	803	15	AA070440	Hepatitis C virus
31	253	73.3	365	17	AA038094	Recombinant oligon
32	253	73.3	552	16	AA079757	DNA fragment HC360
33	253	73.3	573	17	AA016620	Hepatitis C virus
34	253	73.3	807	13	AA020940	Hepatitis C virus
35	253	73.3	1258	13	AA032452	HCV core-envelope
36	253	73.3	1258	13	AA024467	HCV core-envelope
37	253	73.3	1880	13	AA03491	NANB hepatitis vir
38	253	73.3	2187	19	AA03491	Cuticle protein 1
39	253	73.3	2433	17	AA012974	HCV EL construct H
40	253	73.3	2540	14	AA043889	NANB hepatitis vir
41	251.4	72.9	533	19	AA042303	HCV core protein C
42	251.4	72.9	570	13	AA021747	HCV clone HCV JH C
43	251.4	72.9	573	17	AA016615	Hepatitis C virus
44	251.4	72.9	573	24	ABK49784	Hepatitis C virus
45	251.4	72.9	803	15	AA070437	CDNA encoding hepa

ALIGNMENTS

RESULT 1	
ID	AA078089 standard; cDNA, 345 BP.
XX	AA078089:
AC	15-AUG-1995 (first entry)
XX	
DE	Hepatitis C virus core region.
XX	
KW	Hepatitis C virus; HCV; primer: probe; detection; diagnosis;
KM	classification: Immunisation; prophylaxis; serotyping; ss.
OS	Hepatitis C virus.
XX	
FT	Key
FT	1..345
FT	/*tag= a
FT	/product= core polypeptide.
FT	1..342
FT	/*tag= b
PN	mat_peptide
XX	
XX	W09425601-A.
XX	
PD	10-NOV-1994.
XX	
XX	27-APR-1994; 94WO-EP01323.
XX	
PR	27-APR-1993; 93EP-0401099.
PR	05-AUG-1993; 93EP-0402019.
XX	
PA	(INNO-) INNOGENETICS NV SA.

XX	Key	Location/Qualifiers
FH	unSURE	50
FT		/tag= a
FT		/note= "base at this position is designated ? in the specification"
FX		
FN		
XX		
XX		
PD		
XX		
XX		
PE		
PR		
PR		
XX		
PA	(INNO-) INNOENETICS NV.	
PI		
PI	Maertens G, Stuyver L;	
DR	WPI; 1996-251460/25.	
DR	P-PSDB; AAR96550.	
XX		
PT	Hepatitis C virus poly:nucleic acid unique to unidentified sub:type	
PT	used to develop probes and primers for new sub:types and vaccines	
XX	to prevent and treat infection	
PS	Claim 6; Fig 3; 150pp; English.	
CC	The sequences AAT72937-127989 represent novel sequences isolated from	
CC	hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,	
CC	4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,	
CC	3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'	
CC	untranslated region (UR), the Core/EI, NS4 or NS5B regions of the	
CC	genome. This sequence represents nucleotides 1-310 from the HCV type 10a	
CC	isolate NE89.	
CC	The new HCV types were isolated from patients with chronic HCV from the	
CC	Beneux countries, France, Cameroon and Vietnam, because of their	
CC	aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR	
CC	amplified, cloned and genotyped. The 5'UR, Core/EI and NS5B regions were	
CC	sequenced either directly or partially and used to classify the new	
CC	viruses into (sub)types based on comparison with known sequences.	
CC	The sequences can also be used to generate the peptides AAR96424-R96524. The	
CC	sequences can also be used to synthesise probes and primers for the	
CC	detection of HCV in a sample. The polypeptides can be used to detect	
CC	anti-HCV antibodies, for HCV typing or to prevent HCV infections.	
XX		
SO	Sequence 310 BP; 65 A; 99 C; 97 G; 48 T; 1 other:	
Query Match	86.4%; Score 298; DB 17; Length 310;	
Best Local Similarity	99.7%; Pred. No. 2e-71;	
Matches 309; Conservative	0; Mismatches 0; Indels 1; Gaps	
QY	1 ATGAGCACACTCTCTTAACCAACAAGAAAAAACAAAGAACCACCAAC-CGGGCCACAG 59	
Db	1 ATGAGCACACTCTCTTAACCAACAAGAAAAAACAAAGAACCACCAACCGGCCACAG 60	
QY	60 GACGTTAAGTTCACAGCGCGCGTGATCGTTGGTGAATTACGTGCATCACCACAG 119	
Db	61 GACGTTAAGTTCACAGCGCGCGTGATCGTTGGTGAATTACGTGCATCACCACAG 120	
QY	120 GGCCCCAGTTGGGTGTGCGTCACTGTGCATAAAGCTTCCGAGCGGTGCGCAACTCGCACT 179	
Db	121 GGCCCCAGTTGGGTGTGCGTCACTGTGCATAAAGCTTCCGAGCGGTGCGCAACTCGCACT 180	
QY	180 AGGGGCAACCCATCCCGCCAGCGCGCGCAAACGAGGGCAGGCTCTCGGCTCAGCCGGG 239	
Db	181 AGGGGCAACCCATCCCGCCAGCGCGCGCAAACGAGGGCAGGCTCTCGGCTCAGCCGGG 240	
QY	240 TACCTTTGGCCCTTATATGGAATAGAGGCTGGCGGGGACAGGTGGCTCTGTCCCGG 299	
Db	241 TACCTTTGGCCCTTATATGGAATAGAGGCTGGCGGGGACAGGTGGCTCTGTCCCGG 299	

OY 300 CGCGGCTC 309
 |||||||||
 Db 301 CGCGGCTC 310

RESULT 3
 AAT27966
 ID AAT27966 standard; DNA: 652 BP.
 XX
 AC AAT27966:
 XX
 DT 11-MAR-1997 (first entry)
 XX
 DE Hepatitis C virus type 1e isolate CAM1078 bases (-238)-414.
 XX
 KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 XX PCR; primer; probe; antibody; infection; ss.
 XX Hepatitis C virus.
 OS

Key Location/Qualifiers
 FH 239..652
 FT CDS /tag=a
 FT /note="encodes amino acids 1-138 of HCV NS5B protein"
 FT

MO9613590-A2.
 XX
 PD 09-MAY-1996.
 XX
 PE 23-OCT-1995; 95WO-EP04155.
 XX
 PR 28-JUN-1995; 95EP-0870076.
 PR 21-OCT-1994; 94EP-0870166.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Maertens G, Stuyver L;
 XX
 DR WPI: 1996-251460/25.
 DR P-PSDB: AAR6555.
 PT
 PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
 PT - used to develop probes and primers for new sub:types and vaccines
 PT to prevent and treat infection
 XX
 PS Claim 6; Fig 3; 150pp; English.
 XX
 CC The sequences AAT27937-T27989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-l, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/EL, NS4 or NS5B regions of the
 CC genome. This sequence represents nucleotides (-238)-414 from the HCV type
 CC 1e isolate CAM1078.
 CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UR, Core/EL and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences can also be used to generate the peptides AAR6424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
 CC
 SO Sequence 652 BP; 123 A; 202 C; 207 G; 118 T; 2 other;
 Query Match 75.8%; Score 261.6; DB 11; Length 652;
 Best Local Similarity 86.7%; Pred. No. 1.0e-61;
 Matches 299; Conservative 0; Mismatches 45; Indels 1; Gaps 1;
 OY 1 ATGAGCACACTTCTTAACCAAGAAACCAAAAGAAACACCAAC-CGCGCCACAG 59
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 239 ATGAGCACACTTCTTAACCTAAGAAACCAAAAGAAACCAACCAACCGCCGACAG 298
 OY 60 GACGTTAGTTCCTCCAGCGCGGTGAGATCGTTGAGAGTTACGTGACGAG 119
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 299 GACGTCANAGTTCCTCCAGCGCGGTGAGATCGTTGAGAGTTACGTGACGAG 358
 OY 120 GCGCCCGAGTTGGGTGTCGTCGAGTCCGAGCGGTGCGACACCTGCGACGT 179
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 359 GCGCTAGATTGGGTGTCGTCGAGCGCGGTGAGATCGTTGAGAGTTACGTGAG 418
 OY 180 AGCGCCCAACCATCCAGCGCGGTGAGATCGTTGAGAGTTACGTGAGAGTTACGTGAG 239
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 419 AGCGCCCAACCATCCAGCGCGGTGAGATCGTTGAGAGTTACGTGAGAGTTACGTGAG 478
 OY 240 TACCTTGGCCCTTATGAGGATGAGGAGTGGGTCGAGGTCCTGTCCTCCG 299
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 479 TACCTTGGCCCTTATGAGGATGAGGAGTGGGTCGAGGTCCTGTCCTCCG 538
 OY 300 CGCGGCTCCGCTAGTTGGGTGTCGAGCGCGGTGAGATCGTTGAGAGTTACGTGAG 344
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 539 CGCGGCTCCGCTAGTTGGGTGTCGAGCGCGGTGAGATCGTTGAGAGTTACGTGAG 583

RESULT 4
 AAO78098
 ID AAO78098 standard; cDNA: 499 BP.
 XX
 AC AAO78098:
 XX
 DT 15-AUG-1995 (first entry)
 XX
 DE Hepatitis C virus core/EL region.
 XX
 KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
 KW classification; immunisation; prophylaxis; serotyping; ss.
 OS
 OS Hepatitis C virus (isolate CAM600).
 XX
 PN WO9425601-A.
 XX
 PD 10-NOV-1994.
 XX
 PE 27-APR-1994; 94WO-EP01323.
 XX
 PR 27-APR-1993; 93EP-0401099.
 PR 05-AUG-1993; 93EP-0402019.
 XX
 PA (INNO-) INNOGENETICS NV SA.
 XX
 PI Maertens G, Stuyver L;
 XX
 DR WPI: 1994-358277/44.
 DR P-PSDB: AAR63360.
 PT
 PT New polynucleotide sequences from hepatitis C virus - and related
 PT vectors, polypeptide(s) and antibodies, useful for immunisation,
 PT treatment, diagnosis and typing of HCV isolates
 XX
 PS Example 10; Page 214; 404pp; English.
 XX
 CC Compositions comprising at least 5, and pref. 8 or more contiguous
 CC nucleotides selected from an HCV type 3 genomic sequence, more
 CC particularly (i) the region spanning positions 417-957 of the
 CC Core/EL region of HCV subtype 3a; (ii) the region spanning positions
 CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
 CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
 CC region spanning positions 8023-8235 of the NS5 region of the BR36
 CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
 CC sequence, may be used as primers to amplify nucleic acid from an
 CC isolate belonging to a specific genotype, or as a probe for specific
 CC detection/classification of nucleic acid. Polypeptides encoded by
 CC the nucleotides in such compositions may be used for immunisation
 CC against HCV, for the detection of antibodies directed against HCV

100 B; 95 A; 148 C; 161 G; 92 T; 3 other;

[illegible][illegible][illegible]

...AACCTATCCCAAGCGCGCCGAACCGAGGCGAGATCTGGCGCAGCCCGG 240

300 GCGGCTGGTGGGCAGGGTGCCCTGTCCCCT 300

ESITIT 5

C AAZ07651;
X

Hepatitis C virus; HCV; J1; J7; HCV-1; HCV-2

EP939128-A2.

21-DEC-1989; 89US-04561A2

Chapter 7

New Hepatitis C Virus isolates

The invention provides two new isolates of hepatitis C virus (HCV).

Exact match	75.38;	Score 259.8;	DB 20;	Length 552;
Best Local Similarity	81.49;			

[illegible][illegible][illegible][illegible]

300 CGGGCTCTGCCCGTCGTGGGGCCCAATGACCCCCCCCCCCCC 390

AT16643

T
01-OCT-1996 (first entry)

hepatitis; ss.
serotyping; antibody; vaccine;

$$/*tag = a$$

22-FEB-1996.

.000000

(HSSII) CH (HSSII)

(USSH) US DEPT HEALTH & HUMAN SERVICES.
(USSH) US SEC DEPT HEALTH.

XX	22-FEB-1996.	
PD		
XX	15-AUG-1995:	95MO-US10398.
PF		
XX	15-AUG-1994:	94US-0290665.
PR		
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
PA	(USSH) US SEC DEPT HEALTH.	
XX	Bukh J, Miller RH, Purcell RH;	
PI		
XX	WPI: 1996-139709/14.	
DR	P-PSDB: AAF92974.	
XX	DNA and amino acid sequence of HCV envelope 1 and core proteins -	
PT	used to determine HCV genotype and as vaccine against HCV infection	
PS	Claim 3: Page 172: 340pp: English.	
XX		
CC	AAT1610-r16661 are cDNAs encoding a core protein gene of 52 HCV	
CC	isolates. The isolated sequences are useful for the prodn. of primers	
CC	useful for detecting the presence of HCV in a sample, the primers	
CC	are also useful for HCV genotyping. Proteins encoded by the cDNAs	
CC	can be used in vaccines for immunising against HCV infection. The	
CC	proteins may also be used to detect antibodies against HCV in serum,	
CC	saliva, lymphocytes or other mononuclear cells. The antibodies may be	
CC	used in the prevention of HCV infection.	
SQ	Sequence 573 BP: 100 A; 182 C; 174 G; 117 T; 0 other:	
Query Match	74.7%; Score 257.8; DB 17: Length 573:	
Best Local Similarity	86.1%; Pred. No. 1,9e-60;	
Matches 297: Conservative	0; Mismatches 47; Indels 1; Gaps	
OY	1 ATAGCAGACACTCTTAACACCAAGAAGAAAACAAGAAACACCACACCGGCCACH-G 59	
Db	1 ATAGGACCAAAATCTTAAACCTCAAAGAAAAAACAAACAAATACACACACCGGCCCATG 60	
OY	60 GAGCTTAAGTTCCAGCGCGCGGTGAGATCTTGTTGAGATTTCAGTTCACACGCAG 119	
Db	61 GATGTGAATTTCCGGGGGCGGCGGATCTTGCGGAGTTACTCTCTCGCGCAGG 120	
OY	120 GGCCCCAGTTGGGTGTGCGTAGAGGCCAAGACTTCGAGCGGTGCCAACCTGGAGT 179	
Db	121 GGCCCCGGGTGGGTGTGCGCGCACGCTCGGAAGACTTCGAGCGGTACACACTGTGCGC 180	
OY	180 AGCGGCCAACCCATCCCGAGGGCGCGCCGAACCGAGAGGAGGTCCTGGGCTCAGCCGG 239	
Db	181 AGCGCTCAGCTATCCCAAGCGCGCGGTCCGAGGGCAGGTTCTGGGGCTCAGCCCggg 240	
OY	240 TACCTTGCGCCCTATATAGGAGATGAGAGGCTGCGGAGTGGGACAGGTGGCTCTGTGCCG 299	
Db	241 TACCTTGCGCCCTTATAGGCAITAGGGCTGTGGGTGGGCGAGGTGGCTCTGTCCCCC 300	
OY	300 CGGGGCTCTCGCCGCTGCTGGGGGCCCAATACACCCCGGGCGCAG 344	
Db	301 CGCGGTTCCAGGCGCTGTGGGGCCCCAATGATCCCGGCGTAGG 345	
RESULT 8		
AAO70439		
ID	AAO70439 standard: cDNA to mRNA: 803 BP.	
XX	AAO70439:	
AC		
XX	18-MAY-1995 (first entry)	
DF		
XX	Recombinant oligonucleotide US-114 used in detection of HCV.	
DE	Hepatitis C virus: HCV, genotype: detection; identification;	
XX	diagnosis: primer; probe; non-structural protein: ss.	
KM		
XX		

KM	HCV infection; vaccine; ds.
XX	Hepatitis C virus.
OS	
XX	
FH	Location/Qualifiers
Key	91..552
CDS	/tag= a
FT	/note= "the stop codon is not indicated"
FT	
XX	
PN	EP939128-A2.
XX	
PD	01-SEP-1999.
XX	
PF	17-SEP-1990; 90EP-0310149.
XX	
PR	21-DEC-1989; 89US-0456142.
PR	15-SEP-1989; 89US-0408045.
PA	(CHIR) CHIRON CORP.
PA	(OYAA/) OYA A.
PI	Cha T, Han J, Houghton M, Irvine BD, Kolberg JA;
PI	Miyamura T, Saito I, Weiner AJ;
XX	
DR	WPI: 1999-480843/41.
XX	P-FSDB; AA14970.
PT	New Hepatitis C Virus isolates, useful for diagnosis of hepatitis
PT	infections and development of vaccines
PS	
XX	Claim 3; Fig 1; 132pp; English.
CC	The invention provides two new isolates of hepatitis C virus (HCV), J1
CC	and J7. These two isolates comprise nucleotide and amino acid sequences
CC	that are distinct from the HCV isolate HCV-1. The nucleotide sequences
CC	may be used to detect non-A, non-B HCV (NANBH) polynucleotides by
CC	hybridisation for diagnosis of NANBH infections. They may also be used to
CC	screen blood donors, donated blood and blood products for this infection.
CC	The isolates may also be used to isolate other naturally occurring
CC	variants of the virus. The polypeptides may be used as a vaccine for
CC	administration to patients to protect against infection with NANBH. The
CC	present sequence represents the J7 C/E domain coding sequence.
SQ	
Sequence	552 BP; 99 A; 160 C; 177 G; 104 T; 12 other:
Query Match	74.6%; Score 257.4; DB 20; Length 552;
Best local similarity	84.9%; Pred. No. 2,4e-60;
Matches	293; Conservative 5; Mismatches 46; Indels 1; Gaps 1.
OY	1 ATGAGCACACTTCTTTAAACCAAGAAGAAAAACCAGGAACAACCAACC-CGGCCACAG 59
Db	91 ATGACACCAAAATCCTTAACCTCYCAAGAAAACCAAACGTAAACACCAACCGTCGCCACAG 150
OY	60 GACGCTTAGATTGCCAGGCGCGGTCAAGTCGTTGGTGAGTTTACGTGCTAACACGACAG 119
Db	151 GACGTAAAGTTCCCKGCGGCGGTGCAGATCGTGTGAGATTACTGTGCCRCGAGG 210
OY	120 GGCCCCCAGTTGGTGTGCGTCGTCAGTGGCGCAAGACTCCGAGCGGTGCAACCTCGAGT 179
Db	211 GGCCCCCAGTTGGTGTGCGTCGTCGTCAGTGAAGACTTCGAGCGGTGCRCAACCTCGTGA 270
OY	180 AGGCGCCAACCATCCCAAGGCGCGCCGAACCGAAGGGCAGCTCTTGSGCTCAAGCCGGG 239
Db	271 AGGCGCCAACCTATCCCAAGGCTGCGCGGCGCCGAGGACACTTGGGCTCACACCTGGG 330
OY	240 TACCTTTGGCCCTATATGGAATGAGGCTCGGGGTGGGCGAGGCTGCTCTGTCGCCG 299
Db	331 TATCTTTGGCCCTCTATATGGAATGAGGCTGAGGCTGGGGGTGGCGAGATGGCTCTGTCAACC 390
OY	300 CGGCGGCTGCGCCGTGTTGGGCGCCAAATGACCCCGGCGCAG 344
Db	391 CGGCGCTCTGGCTAGTTGGGCGCCYAMTGAACCCCGGCGTAGG 435

	RESULT	11
ID	AAQ11075	standard; DNA: 552 BP.
XX	AAQ11075;	
AC	AAQ11075;	
XX	30-MAY-1991	(first entry)
DT	Fragment of hepatitis C virus J7 isolate C/E domain.	
XX	Hepatitis C virus: HCV-J1; HCV-J7; vaccines; NANBH; ss.	
DE	Hepatitis C virus: HCV-J1; HCV-J7; vaccines; NANBH; ss.	
XX	Hepatitis C virus:	
KW	Hepatitis C virus:	
XX	Hepatitis C virus:	
OS	Hepatitis C virus:	
XX		
FH	Key	Location/Qualifiers
FT	CDS	91..552
FT		/tag= a
FT		/label= HCV-J1 C/E domain
XX		
PN	EP419182-A.	
PD	27-MAR-1991.	
PE	17-SEP-1990;	90EP-0310149.
PR	21-DEC-1989;	89US-0456142.
PR	15-SEP-1989;	89US-0408045.
XX		
PA	(CHIR-) CHIRON CORP.	
PI	Miyamura T, Saito T, Houghton M, Weiner AJ, Han J;	
PI	Kolberg JA, Chata T-A, Irvine BD;	
DR	WPI; 1991-088781/13.	
DR	P-PsDB; AARI1274.	
XX	New isolates J1 and J7 of hepatitis C virus - confg. specified	
PT	DNA and amino acid sequences, used in diagnosis, recombinant	
PT	protein prodn. and vaccine	
XX		
PS	Disclosure; fig 1; 109pp; English.	
XX	This is a fragment of the hepatitis C virus (HCV) J7 isolate C/E	
CC	domain. This is one of the domains of the viral isolate exhibiting	
CC	heterogeneity w.r.t the HCV1 isolate. This sequence has an imp-	
CC	ortant potential use as a probe in diagnostic assays and vaccine	
CC	development. Antibodies directed against it can be used for	
CC	screening antiviral agents and for isolation of non-A-non-B hepat-	
CC	itis (NANBH).	
CC	See also AAQ11076-79.	
CC		
XX	Sequence 552 BP; 98 A; 161 C; 176 G; 103 T; 14 other;	
SQ		
	Query Match	74.5%; Score 257; DB 12; Length 552;
	Best Local Similarity	84.6%; Pred. No. 3e-60;
	Matches 292; Conservative	6; Mismatches 46; Indels 1; Gaps 1.
OY	1 ATGAGCACACTCTTCATACCAACAAAAAAGAAAACCAAACCAACC-CGGCCACAG	59
Db	91 ATGAGCACAACCTTCTTAACCTTSAAGAAAAAACCAACAGTAAACACCAACCGGCCACAG	150
OY	60 GACGTAAAGTTCCCGAGCGCGGCATGTCAGATGTTGGTGAGTTTACGTGTCACCAACGAGG	119
Db	151 GAGGTAAAGTCCCKGCGGTGTCAGATGTCGTGGTAGATTACTTGTGCGCGAGG	210
OY	120 GGCCCCCAGTTGGGTGTGTCGTCAGTGGCGAAGACTTCCAGCGGTGCAACCTCGCAGT	179
Db	211 GGCCCCCAGTTGGGTGTGTCGTCAGTGGCGAAGACTTCCAGCGGTGCAACCTCGCAGT	270
OY	180 AGCGGCAACCATTCGCCAGGCGCGCCGAACCGAAGGCGAGGCTCTGGGCTCAGCCCGG	239
Db	271 AGCGGCAACCATTCGCCAGGCGCGCCGAACCGAAGGCGAGGCTCTGGGCTCAGCCCGG	330

[illegible]

Db	271	AGGCGACAACCTATCCCCAAGGTCCGGCGGCCGAAGGCAGACTTGGC
Qy	240	TACCCTTGCCCCATATAGGAATGAGGCGTCGGGCTGGCAGGTTGGC

[illegible]AC AAQ79756;
XX

DE Hepatitis C virus J7 C/E domain clone c.
XX

	Location/Qualifiers
Key	91..552
mat_peptide	/tag=a /trans_except= pos:355..357, aa:Gly /transl_except= pos:426..428, aa:Ala
XX	
XX	
OS	Hepatitis C virus.
XX	
XX	
KW	nucleic acid C/D/E domain; anti-HCV vaccine development
NM	hepatitis C virus J7 C/E domain; anti-HCV vaccine development
NC	non-A non-B viruses; diagnostic polypeptides; HCV probes; ss.
OS	Hepatitis C virus.
XX	
XX	

13-DEC-1994.

15-SEP-1989; 89US-0408045.

21-DEC-1989;	89US-0456142.
04-JAN-1991;	91US-0637380.

24-FEB-1994; 94US-0201066.
(CHTR) CHIRON CORP

(NAHE-) NAT INST OF HEALTH JAPAN.

WPT; 1995-030306/04.
P-PSDB: AAR66615
Cuda T, Han J, Houghton M, Irvine BD, Kolberg JA;
Miyamura T, Saito I, Weiner AJ;

Method of detecting bacteria:

Claim 1; Fig 1; 45pp; English

AA079756 encodes AAR66615 the prod. of the hepatitis C virus (HCV).
J7 C/E domain clone C sequence. Shows a high degree of homology to the

recombinant protein prodn. and anti-HCV vaccine development.

Sequence 552 BP; 100 A; 163 C; 180 G; 109 T; 0 other.

PR 08-MAY-1995; 95US-0436965.

XX
XX (CHIR) CHIRON CORP.
PA (NAHE-) NAT INST OF HEALTH JAPAN.
XX
XX

PI Miyamura T, Saito I;

XX WPI, 1999-16619/14.

DR P-PSDB; AAY01609.

PT Immunassays for Asiatic strains of hepatitis C virus - for
XX diagnosis of infection and screening blood supplies

PS Disclosure: Fig 1, 43pp: English.

XX
XX The present sequence represents the consensus sequence of the coding
CC strand of a new hepatitis C virus (HCV), J7, C/E domain. The J7 and J1
CC (also a new HCV isolate) isolates comprise sequences which are distinct
CC from the prototype HCV isolates, HCV1. The specification describes
CC immunassays for HCV based on antigens from Asiatic strains not
CC cross-reactive with HCV-1. The assays are used for diagnosis of HCV
CC infection and to screen donated blood. The anti-HCV antibodies are also
CC useful therapeutically and prophylactically (passive immunisation); in
CC screening for antiviral agents; for isolation, purification and
CC identification of non-A, non-B hepatitis virus (e.g. by affinity
CC chromatography) and to raise anti-idiotypic antibodies (useful for
CC treatment or diagnosis and to determine immunogenic regions of the
CC HCV antigens).

SQ Sequence 552 BP; 100 A; 163 C; 180 G; 109 T; 0 other;

Query Match 74.3%; Score 256.2; DB 20; Length 552;
Best Local Similarity 85.8%; Pred. No. 5e-60;

Matches 296; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 1 ATGAGCAGACTTCTTAACCAAGAAAAACCAAGAACCAACCC-CGGCCACAG 59
DB 91 ATGAGCAGCAATCTTAACCCCAAGAAAAACCAAGAACCAACCCGTGCCACAG 150
QY 60 GACGTTAAGTCCAGGCGGCGTCAATCGTTGGTGAATTTACGCTACCAACGAG 119
DB 151 GACGTTAAGTCCCGGCGGCGTCAATCGTTGGTGAATTTACGCTACCAACGAG 210
QY 120 GGGCCCCAGTGGTGTGCTGCACTGCGCAAGACTTCCGAGCGGTGCAACTGCGAGT 179
DB 211 GGGCCCCAGTGGTGTGCTGCACTGCGCAAGACTTCCGAGCGGTGCAACTGCGAGT 270
QY 180 AGGCGCAACCATCCCGAGGCGCGCAAGCGGAGGAGTCTGGCTCAGCCCGGG 239
DB 271 AGGCGCAACCATCCCGAGGCGCGCAAGCGGAGGAGTCTGGCTCAGCCCGGG 330
QY 240 TACCTTGGCCCTATATGGAATGAGGCTGCGGTTGGCAGGATGCTCTGCTCCCG 299
DB 331 TATCTTGGCCCTATATGGAATGAGGCTGCGGTTGGCAGGATGCTCTGCTCCCG 390
QY 300 CGCGGCTCTCGCCCTGCTGGGCGCAATATGACCCCGCGCAGG 344
DB 391 CGCGGCTCTCGCCCTGCTGGGCGCAATATGACCCCGCGCAGG 435

Search completed: February 19, 2003, 00:28:01
Job time : 244 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 00:13:35 : Search time 1463 Seconds
(Without alignments)
3819.168 Million cell updates/sec

Title: US-09-873-224-147
345
Sequence: 1 ATGAGCACACTTCTTAAC.....AAATGACCCCGCGCAGAGA 345

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

EST:*

- 1: em_estba:*
- 2: em_esthm:*
- 3: em_estin:*
- 4: em_estmv:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_iny:*
- 20: em_gss_pln:*
- 21: em_gss_vit:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	15.4	925	17	CNS0091P
2	52.8	15.3	879	17	CNS0200G
3	51.8	13.0	844	17	CNS0052P
4	46.6	13.5	1523	14	BO948210
5	45.4	13.2	925	17	CNS0091P
6	45	13.0	844	17	CNS0052P

7	43.8	12.7	914	17	CNS00C2P	AL059740 Drosophila
8	43.8	12.7	932	17	CNS00720	AL066742 Drosophila
9	43.4	12.6	712	17	AG076938	AG076938 Pan trogl
10	43.2	12.5	645	17	CNS01213	AL101589 Drosophila
11	43.2	12.5	932	17	CNS00720	AL066742 Drosophila
12	43.2	12.5	1057	17	CNS046B0	AL276477 Tetraodon
13	42.8	12.4	442	9	AF367693	AF367693
14	42.8	12.4	1101	17	CNS0160E	AL107216 Drosophila
15	41.8	12.1	1201	17	CNS016BR	AL106545 Drosophila
16	41.6	12.1	1201	17	CNS016BR	AL106528 Drosophila
17	41.4	12.0	926	17	AG060188	AG060188 Pan trogl
18	41.2	11.9	1009	17	CNS010EM	AL098882 Drosophila
19	41.2	11.9	1009	17	CNS010EM	AL098882 Drosophila
20	40.8	11.8	384	17	CNS043C6	AL272751 Tetraodon
21	40.8	11.8	784	17	AG060328	AG060328 Pan trogl
22	40.8	11.8	961	17	CNS01087	AL098641 Drosophila
23	40.8	11.8	1103	13	BM346874	BM346874 AGENCOURT
24	40.8	11.8	1201	17	CNS014B0	BM802578 AGENCOURT
25	40.2	11.7	1191	14	BM802578	AL106711 Drosophila
26	39.4	11.4	787	17	AG060334	AL098641 Drosophila
27	39.4	11.4	961	17	CNS01087	AL106711 Drosophila
28	39.4	11.4	1201	17	CNS01087	AL066051 Drosophila
29	39.2	11.4	767	17	AG076679	AG076679 Pan trogl
30	39.2	11.3	935	17	CNS006XK	AL066051 Drosophila
31	39.2	11.3	948	14	BO646304	BO646304 AGENCOURT
32	39.2	11.3	1038	17	AG060254	AG060254 Pan trogl
33	39.2	11.3	1293	17	AG064099	AG064099 Pan trogl
34	38.8	11.2	692	17	CNS007WH	AA024468 ze74g10.r
35	38.8	11.2	692	17	CNS015W3	AL105923 Drosophila
36	38.8	11.2	1201	17	CNS016BR	AL105981 Drosophila
37	38.8	11.2	1201	17	CNS016BR	AL106054 Drosophila
38	38.8	11.2	1201	17	CNS015V4	AL106054 Drosophila
39	38.6	11.2	538	9	AL514267	AL514267 AL514267
40	38.6	11.2	846	17	CNS010R0	AL099337 Drosophila
41	38.6	11.2	910	17	CNS0060N	AL065629 Drosophila
42	38.6	11.2	961	17	AG080470	AG080470 Pan trogl
43	38.6	11.2	1180	14	BO422969	BO422969 AGENCOURT
44	38.4	11.1	255	12	BM687972	BM687972 335326 BA
45	38.4	11.1	465	12	BM690214	BM690214 338670 BA

ALIGNMENTS

RESULT 1
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACRI9D16 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammosier in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial

COMMENT

source

1. .925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR1916"
/note="end : TET3"
BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

Query Match
Best Local Similarity 13.2%; Score 45.4; DB 17; Length 925;
Matches 37; Conservative 142; Mismatches 103; Indels 1; Gaps 1;

QY 42 CACCAACCCCGCCACGACGTTAGTCCAGCGCGGCTGCTGTTGGTGGAGT 101
DB 842 MABCCSSSSSSCCGASARGVKVASGAGRGAGGAGSASISSAACBSSSSCSAS 783
QY 102 TACTGCTACACGACGAGGCGCCCGACGTTGGTGTGCTGCTGCTGCTGCTG 783
DB 782 CWSASS 783
QY 162 CGGTGCAACCTGCGAGAGCGCCACACCCATCCCGAGCGCGCCGCAACCGAGCG 724
DB 723 CTSVASS 724
QY 222 TCTGAGGCTACGCGCGGCTGACCTTGGCCCTATATGGGAATGAGGCTGCGGTG 664
DB 663 SSSGSS 664
QY 282 GGGTGGCTCTCTCCCGCGCGCTCTGCGCGCTGCGGAGCG 324
DB 603 SSGGSS 604

RESULT 6
LOCUS
DEFINITION

CNS0052P 844 bp DNA linear GSS 03-JUN-1999
BACR1916 of RPT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL056652.1 GI:4932342
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Ephidrotidea; Drosophilidae; Diptera; Brachycera; Muscomorpha;
Genoscope.
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
Aeron Mammoser BAC library was prepared by Kazuhiro Osoegawa
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPT-98 and was constructed by partial
digestion of Drosophila DNA provided by the BDGP from the
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library,
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm
1. 844
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

FEATURES
source

BASE COUNT 261 a 112 c 92 g 35 t 344 others
ORIGIN

Query Match
Best Local Similarity 13.0%; Score 45; DB 17; Length 844;
Matches 63; Conservative 115; Mismatches 129; Indels 0; Gaps 0;

QY 16 AAACCAAAAGAAAAACAAAGAAACACACCCCGCCAGCAGCAGTGTGCTCCAG 75
DB 339 AAAAACAACAAAAAACAASASVSSASVAGSSMAAGASCBGCAAAAMKACGSS 0;
QY 76 GCGGCGCTCAGTCTGTTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 398
DB 399 SSSASRSMSS 398
QY 136 TGCGTCACTGCGCAAGACTTCCAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTG 135
DB 459 SSGGVSSVSGGAVRGRCMCCSMVCCCMSCMCMSCMCMSCMCMSCMCMSCMCMSC 458
QY 196 CAGGCGCGCGCGCAACCGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 195
DB 519 SVGGVGGSSRRAGRSRGGGSSSVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 518
QY 256 ATGGGAATGAGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 255
DB 579 SSGAVSRGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 578
QY 316 CGTGGG 322
DB 639 RGGGGG 645

RESULT 7
LOCUS
DEFINITION

CNS00CZP 914 bp DNA linear GSS 04-JUN-1999
BACR26P05 of RPT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL059740
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Ephidrotidea; Drosophilidae; Diptera; Brachycera; Muscomorpha;
Genoscope.
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
Aeron Mammoser BAC library was prepared by Kazuhiro Osoegawa
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPT-98 and was constructed by partial
digestion of Drosophila DNA provided by the BDGP from the
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library,
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm
1. 914
Location/Qualifiers
/organism="Drosophila melanogaster"

FEATURES
source

Db 713 CCGCGGSCGSGGSCGCCSCCCSCCCSCCGSSSGGSGGSGGCG 772

QY 196 CAGGGGCGGCCAACCAGAGGTCTGTGAGCTACGCCGGTACCTTGCCCTAT 255

Db 773 SCGGSCGCGCGGCGGCGGSSGCGSSGCGSSGCGGSSGCGSSGCGC 832

QY 256 ATGGGATGATGAGGCTGCGGCGAGGCTGCTGCTGCGGCGGCGCTGCGCCCT 315

Db 833 SSSGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 892

QY 316 CGTGGGCGCC 325

Db 893 CCGGCGGCGSS 902

RESULT 12
CNS046BO/c
LOCUS
DEFINITION
CNS046BO 1057 bp DNA linear GSS 18-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
085G14 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION
AL276477.1 GI:8010680

VERSION
AL276477.1

KEYWORDS
GSS: genome survey sequence.

SOURCE
Tetraodon nigroviridis.

ORGANISM
Tetraodon nigroviridis

REFERENCE
Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Barnot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
Sautin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished

JOURNAL
REFERENCE
2 (bases 1 to 1057)

AUTHORS
Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Barnot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
Sautin, W., and Weissenbach, J.

TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished

JOURNAL
REFERENCE
3 (bases 1 to 1057)

AUTHORS
Weissenbach, J.

COMMENT
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
source
1..1057
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="085G14"
/note="Genoscope sequence ID : C08G085BD07LP1-end : T7"

BASE COUNT
154 a 432 c 262 g 138 t 71 others

ORIGIN
Query Match 12.5%; Score 43.2; DB 17; Length 1057;
Best Local Similarity 44.6%; Pred. No. 4.8; Mismatches 110; Indels 1; Gaps 1;
Matches 107; Conservative 22;

QY 104 CGTGCTACAGCGAGGCGCCCGAGTGTGCTGCTGAGTGGCGAAGACTTCCAGGCG 163

Db 933 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 874

QY 164 GTGCAACCTGCGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 223

Db 873 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 814

QY 224 CTGGGCTACGCGCGGAGTACCTTGAGCCCTATATAGGATAGAGGCTGCGGAGG 283

Db 813 SSSGCG 755

QY 284 GTGGCTCTGCTCCCGCGCGGCTGCTGCGGCGGCTGCTGCGGCGGCGGCGGCGG 343

Db 754 GCG 695

RESULT 13
AF367693/c
LOCUS
DEFINITION
AF367693 Homo sapiens differential display (Yong LL) Homo sapiens
cDNA clone dominant regulation-3, mRNA sequence.

ACCESSION
AF367693.1 GI:14089399

VERSION
AF367693.1

KEYWORDS
EST.

SOURCE
human.

ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 442)

AUTHORS
Yong, L.L.

TITLE
Dominant regulation

JOURNAL
Unpublished (2001)

COMMENT
Contact: Yong LL
Transplantation
Shanghai Clinical Medicine School, Pudon University
Wujin Road 85, Shanghai, 200080, China
Email: koucai@online.sh.cn

FEATURES
source
1..442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="dominant regulation-3"
/note="Homo sapiens differential display (Yong LL)"
/note="isolated by differential display"

BASE COUNT
172 a 121 c 86 g 63 t

ORIGIN
Query Match 12.4%; Score 42.8; DB 9; Length 442;
Best Local Similarity 63.7%; Pred. No. 7.1;
Matches 65; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 227 GCGTCAGCGCGGCTACCTTGCGCCCTATATAGGATAGAGGCTGCGGAGGCGG 286

Db 113 GCGTCAGGATGCTCACTATAAAAAAATAAATACGAGGTATGGGTGCGGTGG 54

QY 287 GCTCCTGTCCCGCGCGGCTGCTGCGGCGGCTGCTGCGGCGGCGGCGGCGG 328

Db 53 GCTGCTGTCTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12

RESULT 14
CNS016UE
LOCUS
DEFINITION
CNS016UE 1101 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN1616 of Drosophila library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION
AL107216.1 GI:5626210

VERSION
AL107216

KEYWORDS
GSS.

SOURCE
Drosophila melanogaster.

ORGANISM
Drosophila melanogaster

REFERENCE
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

AUTHORS
1 (bases 1 to 1101)

TITLE
Direct Submission

JOURNAL
Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage ;
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

Thu Feb 20 14:41:31 2003

us-09-873-224-147.rn1

Page 1

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 00:19:05 ; Search time 77 seconds
(without alignments)
1374.071 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 345

Sequence: 1 ATGACGACACTCTCTAAACC.....AAATGACCCCGCGCAGGA 345

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCUTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	309	89.6	309	US-08-836-075A-49	Sequence 49, Appl
2	261.6	75.8	652	US-08-836-075A-59	Sequence 59, Appl
3	259.4	75.2	573	US-08-290-665A-136	Sequence 136, App
4	259.4	75.2	573	PCT-US95-10398-136	Sequence 136, App
5	257.8	74.7	573	US-08-290-665A-141	Sequence 141, App
6	257.8	74.7	573	PCT-US95-10398-141	Sequence 141, App
7	257.8	74.7	803	US-08-157-235-4	Sequence 4, Appl
8	257.8	74.7	803	US-08-157-235-5	Sequence 5, Appl
9	256.2	74.3	573	US-08-290-665A-135	Sequence 135, App
10	256.2	74.3	573	US-08-290-665A-137	Sequence 137, App
11	256.2	74.3	573	US-08-290-665A-138	Sequence 138, App
12	256.2	74.3	573	PCT-US95-10398-137	Sequence 137, App
13	256.2	74.3	573	PCT-US95-10398-138	Sequence 138, App
14	256.2	74.3	1037	US-08-462-195-1	Sequence 1, Appl
15	256.2	74.3	1037	US-08-462-195-1	Sequence 1, Appl
16	256.2	74.3	1037	US-08-636-883-1	Sequence 1, Appl
17	256.2	74.3	1037	US-09-127-829-1	Sequence 1, Appl
18	254.6	73.8	573	US-08-290-665A-107	Sequence 107, App
19	254.6	73.8	573	US-08-290-665A-114	Sequence 114, App
20	254.6	73.8	573	US-08-290-665A-107	Sequence 107, App
21	254.6	73.8	573	PCT-US95-10398-110	Sequence 110, App
22	254.6	73.8	573	PCT-US95-10398-114	Sequence 114, App
23	254.6	73.8	573	PCT-US95-10398-119	Sequence 119, App
24	254.6	73.8	803	US-08-157-235-2	Sequence 2, Appl
25	253.6	73.5	573	US-08-290-665A-139	Sequence 139, App
26	253.6	73.5	573	PCT-US95-10398-139	Sequence 139, App
27	253.6	73.5	803	US-08-157-235-6	Sequence 6, Appl

28	253	73.3	573	US-08-290-665A-113	Sequence 113, App
29	253	73.3	573	PCT-US95-10398-113	Sequence 113, App
30	253	73.3	803	US-08-157-235-1	Sequence 1, Appl
31	253	73.3	1539	US-08-470-426B-17	Sequence 17, Appl
32	253	73.3	1863	US-08-470-426B-14	Sequence 14, Appl
33	253	73.3	2433	US-08-612-973-49	Sequence 49, Appl
34	253	73.3	2433	US-08-927-597-49	Sequence 49, Appl
35	251.4	72.9	345	US-08-324-977-7	Sequence 7, Appl
36	251.4	72.9	345	US-08-384-616-7	Sequence 7, Appl
37	251.4	72.9	345	US-08-904-686A-7	Sequence 7, Appl
38	251.4	72.9	573	US-08-315-850-7	Sequence 7, Appl
39	251.4	72.9	573	US-08-290-665A-108	Sequence 108, App
40	251.4	72.9	573	PCT-US95-10398-108	Sequence 108, App
41	251.4	72.9	803	US-08-157-235-3	Sequence 3, Appl
42	251.4	72.9	1167	US-08-324-977-9	Sequence 9, Appl
43	251.4	72.9	1167	US-08-384-616-9	Sequence 9, Appl
44	251.4	72.9	1167	US-08-904-686A-9	Sequence 9, Appl
45	251.4	72.9	1167	US-09-315-850-5	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-836-075A-49
Sequence 49, Application US/08836075A
Patent No. 6180768
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
APPLICANT: STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
TITLE OF INVENTION: AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 309 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTISENSE: NO
US-08-836-075A-49

Query Match
Best Local Similarity 89.6%; Score 309; DB 4; Length 309;
Matches 309; Conservative 100.0%; Pred. No. 5.9e-77;
Mismatches 0; Indels 0; Gaps 0;

1 ATGAGCAGCTTCTTAACCAAGAAAAACCAAGAAACCAACCCGCGCAGAG 60
1 ATGAGCAGCTTCTTAACCAAGAAAAACCAAGAAACCAACCCGCGCAGAG 60
61 ACCTTAAGTCCCAAGCGCGGTCAGATCTGTGAGTTACGCTACACAGCAGG 120
61 ACCTTAAGTCCCAAGCGCGGTCAGATCTGTGAGTTACGCTACACAGCAGG 120
121 GCCCCAGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 120
121 GCCCCAGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 120
181 GCGCGCAACCCATCCCAAGCGCGGTCAGATCTGTGAGTTACGCTACACAG 180
181 GCGCGCAACCCATCCCAAGCGCGGTCAGATCTGTGAGTTACGCTACACAG 180
241 ACCCTTGAGCCCTTATGGAATGAGGCTGGGCTGGGCTGGGCTGGGCTGGG 240
241 ACCCTTGAGCCCTTATGGAATGAGGCTGGGCTGGGCTGGGCTGGGCTGGG 240
301 GCGGCTCTC 309
301 GCGGCTCTC 309

RESULT 2

US-08-836-075A-59
Sequence 59, Application US/08836075A
Patent No. 6180768

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
APPLICANT: STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836.075A
FILING DATE: 21 Apr 1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995

ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOHETICAL: NO
ANTI-SENSE: NO
US-08-836-075A-59

Query Match
Best Local Similarity 75.8%; Score 261.6; DB 4; Length 652;
Matches 299; Conservative 86.7%; Pred. No. 9.9e-64;
Mismatches 45; Indels 1; Gaps 1;

1 ATGAGCAGCTTCTTAACCAAGAAAAACCAAGAAACCAACCCGCGCAGAG 59
239 ATGAGCAGCTTCTTAACCAAGAAAAACCAAGAAACCAACCCGCGCAGAG 298
60 GAGCTTAAGTCCCAAGCGCGGTCAGATCTGTGAGTTACGCTACACAGCAG 119
299 GAGCTTAAGTCCCAAGCGCGGTCAGATCTGTGAGTTACGCTACACAGCAG 358
120 GCCCCAGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 358
120 GCCCCAGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 358
359 GAGCTTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 358
180 AGCGCCCAACCCATCCCAAGCGCGGTCAGATCTGTGAGTTACGCTACACAG 239
419 AGCGCCCAACCCATCCCAAGCGCGGTCAGATCTGTGAGTTACGCTACACAG 478
240 TACCTTGAGCCCTTATGGAATGAGGCTGGGCTGGGCTGGGCTGGGCTGGG 478
479 TACCTTGAGCCCTTATGGAATGAGGCTGGGCTGGGCTGGGCTGGGCTGGG 478
300 GCGGCTCTC 309
539 GCGGCTCTC 309

RESULT 3

US-08-290-665A-136
Sequence 136, Application US/08290665A
Patent No. 5882852

GENERAL INFORMATION:

APPLICANT: BORK, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290.665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 136:
SEQUENCE CHARACTERISTICS:
LENGTH: 652 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

```

SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S52
US-08-290-665A-136

```

```

Query Match          75.2%  Score 259.4; DB 2; Length 573;
Best Local Similarity 86.4%; Pred. No. 3.9e-63;
Matches 298; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

```

```

QY 1 ATGAGCACACTTCTTAACACAGAAAGAAACCAAGAACACACACAC-CCGGCCACAG 59
DB 1 ATGAGCACACTTCTTAACACAGAAAGAAAGAAACCAAGAACACACACCGGCCACAG 60
QY 60 GACGTTAAGTTCACAGCGCGGTGATGTTGATGAGTTTACGTTACACAGCAGG 119
DB 61 GACGTTAAGTTCACAGCGCGGTGATGTTGATGAGTTTACGTTGTTGCGCGCAGG 120
QY 120 GGGCCCCAGTTGGGTGTCGTCAGTCGCGCAAGACTTCCGAGCGGTCCGACCTGCGAGT 179
DB 121 GGGCCACGATTGGGTGTCGTCGCGCAAGACTTCTGAACGCTGACAGCCTGCGCGA 180
QY 180 AGCGGCCAACCCATCCCGAGGGCGCGGACCGAGGCGAGGTCTCGGTGAGCGCGGG 239
DB 181 CGAGCAGACCTTATCCCAAGGGGCGTGGAGCGAAGGCGGTCTCGGTGAGCGCGGG 240
QY 240 TACCTTGGCCCCATATATGGAATGAGGGCTCGGGTGGGAGGTGAGTCTGTCGCCG 299
DB 241 TACCTTGGCCCCATATATGGAATGAGGGCTCGGGTGGGAGGTGAGTCTGTCGCCA 300
QY 300 CGCGGCTCGCCCGTGTGCGGCGCCAAATGACCCCGCGCGAG 344
DB 301 CGCGGCTCGCCCGTGTGCGGCGCCAAAGACCCCGCGCGAG 345

```

```

RESULT 4
PCT-US95-10398-136
Sequence 136, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290,665
FILING DATE: 15 AUGUST 1994

```

```

ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 136:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S52
PCT-US95-10398-136

```

```

Query Match          75.2%  Score 259.4; DB 5; Length 573;
Best Local Similarity 86.4%; Pred. No. 3.9e-63;
Matches 298; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

```

```

QY 1 ATGAGCACACTTCTTAACACAGAAAGAAAGAAACCAAGAACACACAC-CCGGCCACAG 59
DB 1 ATGAGCACACTTCTTAACACAGAAAGAAAGAAAGAAACCAAGAACACACACCGGCCACAG 60
QY 60 GACGTTAAGTTCACAGCGCGGTGATGTTGATGAGTTTACGTTTACACAGCAGG 119
DB 61 GACGTTAAGTTCACAGCGCGGTGATGTTGATGAGTTTACGTTTACACAGCAGG 120
QY 120 GGGCCCCAGTTGGGTGTCGTCAGTCGCGCAAGACTTCCGAGCGGTCCGACCTGCGAGT 179
DB 121 GGGCCACGATTGGGTGTCGTCGCGCAAGACTTCTGAACGCTGACAGCCTGCGCGA 180
QY 180 AGCGGCCAACCCATCCCGAGGGCGCGGACCGAGGCGAGGTCTCGGTGAGCGCGGG 239
DB 181 CGAGCAGACCTTATCCCAAGGGGCGTGGAGCGAAGGCGGTCTCGGTGAGCGCGGG 240
QY 240 TACCTTGGCCCCATATATGGAATGAGGGCTCGGGTGGGAGGTGAGTCTGTCGCCG 299
DB 241 TACCTTGGCCCCATATATGGAATGAGGGCTCGGGTGGGAGGTGAGTCTGTCGCCA 300
QY 300 CGCGGCTCGCCCGTGTGCGGCGCCAAATGACCCCGCGCGAG 344
DB 301 CGCGGCTCGCCCGTGTGCGGCGCCAAAGACCCCGCGCGAG 345

```

```

RESULT 5
US-08-290-665A-141
Sequence 141, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 141:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 21
US-08-290-665A-141

```

```

Query Match      74.7%; Score 257.8; DB 2; Length 573;
Best Local Similarity 86.1%; Pred. No. 1.1e-62;
Matches 297; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

```

```

QY 1 ATGAGCACACTTCTTAACCAAGAAACCAAAAGAACCAACCCCGCCACA-G 59
Db 1 ATGAGCACAAATCTTAACCTTAAGAAACCAACCAACCAACCAACCTGCGCATG 60
QY 60 GACGTTAAGTCCAGAGCGCGGTACATGTTGGTGAAGTTAAGTGTACACAGCAAG 119
Db 61 GATGTAATTCCTCCAGCGCGGCGCCAGATGTTGGCGAGTTACTGTGCGCGGCAAG 120
QY 120 GCGCCCAAGTTGGGTGTGCGTGAAGTGGCAAGACTTCCGACGGTCCGACCTTCGCACT 179
Db 121 GCGCCCGGTTGGGTGTGCGTGAAGTGGCAAGACTTCCGACGGTCCGACCTTCGCACT 180
QY 180 AGCGGCCAACCCATCCCGAGGCGCGCCGACCAAGGAGGAGTCTGGGCTCAGCGCGG 239
Db 181 AGCGGTACAGCTATCCCAAGGCGCGCGGTCCGAGGAGGAGTCTGGGCTCAGCGCGG 240
QY 240 TACCTTGGCCCTATATATGGAATGAGGCTCGGGTGGGAGGAGTCTGCTGCTCCCG 299
Db 241 TACCTTGGCCCTTACGCAATGAGGCTGTGGGTGGGAGGAGTCTGCTGCTCCCG 300
QY 300 CGCGGCTTCGCGCGTGTGGGCGCCCAATGACCCCGCGCAGG 344
Db 301 CGCGGTTCCAGGCGCTTGTGGGCGCCCAATGATCCCGCGCTAGG 345

```

```

RESULT 6
PCT-US95-10398-141
Sequence 141, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK

```

```

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 141:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 21
PCT-US95-10398-141

```

```

Query Match      74.7%; Score 257.8; DB 5; Length 573;
Best Local Similarity 86.1%; Pred. No. 1.1e-62;
Matches 297; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

```

```

QY 1 ATGAGCACACTTCTTAACCAAGAAACCAAAAGAACCAACCCCGCCACA-G 59
Db 1 ATGAGCACAAATCTTAACCTTAAGAAACCAACCAACCAACCAACCTGCGCATG 60
QY 60 GACGTTAAGTCCAGAGCGCGGTACATGTTGGTGAAGTTAAGTGTACACAGCAAG 119
Db 61 GATGTAATTCCTCCAGCGCGGCGCCAGATGTTGGCGAGTTACTGTGCGCGGCAAG 120
QY 120 GCGCCCAAGTTGGGTGTGCGTGAAGTGGCAAGACTTCCGACGGTCCGACCTTCGCACT 179
Db 121 GCGCCCGGTTGGGTGTGCGTGAAGTGGCAAGACTTCCGACGGTCCGACCTTCGCACT 180
QY 180 AGCGGCCAACCCATCCCGAGGCGCGCCGACCAAGGAGGAGTCTGGGCTCAGCGCGG 239
Db 181 AGCGGTACAGCTATCCCAAGGCGCGCGGTCCGAGGAGGAGTCTGGGCTCAGCGCGG 240
QY 240 TACCTTGGCCCTATATGGAATGAGGCTCGGGTGGGAGGAGTCTGCTGCTCCCG 299
Db 241 TACCTTGGCCCTTACGCAATGAGGCTGTGGGTGGGAGGAGTCTGCTGCTCCCG 300
QY 300 CGCGGCTTCGCGCGTGTGGGCGCCCAATGACCCCGCGCAGG 344
Db 301 CGCGGTTCCAGGCGCTTGTGGGCGCCCAATGATCCCGCGCTAGG 345

```

```

RESULT 7
US-08-157-235-4
Sequence 4, Application US/08157235
Patent No. 555016
GENERAL INFORMATION:
APPLICANT: OKAMOTO, Hiroaki
TITLE OF INVENTION: OLIGONUCLEOTIDES OF HCV, PRIMERS AND
TITLE OF INVENTION: PROBES THEREFROM, METHOD OF DETERMINING HCV GENOTYPES,
TITLE OF INVENTION: AND METHOD OF DETECTING HCV IN SAMPLES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Beveridge, Degrandi, Wellacher & Young

```

Thu Feb 20 14:41:31 2003

us-09-873-224-147.rml

Page 5

STREET: 1850 M Street N.W., Suite 800
CITY: Washington
STATE: D.C.
COUNTRY: D.C.
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/157,235
FILING DATE: 24-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 354370/92
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Robert G. Wellacher
REGISTRATION NUMBER: 20,531
REFERENCE/DOCKET NUMBER: 06/87-49206
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-659-2811
TELEFAX: 202-659-1462
TELEX: 64470
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 803 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-157-235-4

Query Match 74.7%; Score 257.8; DB 1; Length 803;
Best Local Similarity 86.1%; Pred. No. 1.2e-62;
Matches 297; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

QY 1 ATGAGCAGACTTCTTAACCAACCAAAAGAAAGAAACCAACACC-CGGCCACAG 59
DB 298 ATGAGCAGACTTCTTAACCTCAAGCAAAAGAAAGAAACCAACACC 357
QY 60 GAGCTAAGTTCCAGAGCGCGGTGAGATGCTGTGAGATTACGTACACGAG 119
DB 358 GAGCTAAGTTCCAGAGCGCGGTGAGATGCTGTGAGATTACGTGTGCGGAGG 417
QY 120 GGGCCCGAGTTGGGTGTGCTGAGTGGCAGAGACTTCCGAGCGTGCACACTCGCAGT 179
DB 418 GGGCCCGAGTTGGGTGTGCTGAGTGGCAGAGACTTCCGAGCGTGCACACTCGCAGG 477
QY 180 AGGCGCCAGACTTCCAGAGCGCGGTGAGATGCTGTGAGATTACGTACACGAG 239
DB 478 CGAGCAGAGCTTATCCCAAGCGCGGTGAGATGCTGTGAGATTACGTGTGCGGAG 537
QY 240 TACCTTGGCCCTATATGAGGATGAGGCTGCGGAGTGGCAGAGGCTCTCTCTCCCG 299
DB 538 TACCTTGGCCCTATATGAGGATGAGGCTGCGGAGTGGCAGAGGCTCTCTCTCCCA 597
QY 300 CGCGGCTCTCGCGCTGCTGGGCGCCAAATGACCCCGGCGGAGG 344
DB 598 CGCGGCTCTCGCGCTGCTGGGCGCCAAATGACCCCGGCGGAGG 642

RESULT 8
US-08-157-235-5
Sequence 5, Application US/08157235
Patent No. 5550016
GENERAL INFORMATION:
APPLICANT: OKAMOTO, Hiroaki
TITLE OF INVENTION: OLIGONUCLEOTIDES OF HCV, PRIMERS AND
TITLE OF INVENTION: PROBES THEREFROM, METHOD OF DETERMINING HCV GENOTYPES,
TITLE OF INVENTION: AND METHOD OF DETECTING HCV IN SAMPLES
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

ADDRESSEE: Beveridge, Degrandi, Wellacher & Young
STREET: 1850 M Street N.W., Suite 800
CITY: Washington
STATE: D.C.
COUNTRY: D.C.
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/157,235
FILING DATE: 24-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 354370/92
FILING DATE: 27-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Robert G. Wellacher
REGISTRATION NUMBER: 20,531
REFERENCE/DOCKET NUMBER: 06/87-49206
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-659-2811
TELEFAX: 202-659-1462
TELEX: 64470
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 803 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-157-235-5

Query Match 74.7%; Score 257.8; DB 1; Length 803;
Best Local Similarity 86.1%; Pred. No. 1.2e-62;
Matches 297; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

QY 1 ATGAGCAGACTTCTTAACCAACCAAAAGAAAGAAACCAACACC-CGGCCACAG 59
DB 298 ATGAGCAGACTTCTTAACCTCAAGCAAAAGAAAGAAACCAACACC 357
QY 60 GAGCTAAGTTCCAGAGCGCGGTGAGATGCTGTGAGATTACGTACACGAG 119
DB 358 GAGCTAAGTTCCAGAGCGCGGTGAGATGCTGTGAGATTACGTGTGCGGAGG 417
QY 120 GGGCCCGAGTTGGGTGTGCTGAGTGGCAGAGACTTCCGAGCGTGCACACTCGCAGT 179
DB 418 GGGCCCGAGTTGGGTGTGCTGAGTGGCAGAGACTTCCGAGCGTGCACACTCGCAGG 477
QY 180 AGGCGCCAGACTTCCAGAGCGCGGTGAGATGCTGTGAGATTACGTACACGAG 239
DB 478 CGAGCAGAGCTTATCCCAAGCGCGGTGAGATGCTGTGAGATTACGTGTGCGGAG 537
QY 240 TACCTTGGCCCTATATGAGGATGAGGCTGCGGAGTGGCAGAGGCTCTCTCTCCCG 299
DB 538 TACCTTGGCCCTATATGAGGATGAGGCTGCGGAGTGGCAGAGGCTCTCTCTCCCG 597
QY 300 CGCGGCTCTCGCGCTGCTGGGCGCCAAATGACCCCGGCGGAGG 344
DB 598 CGCGGCTCTCGCGCTGCTGGGCGCCAAATGACCCCGGCGGAGG 642

RESULT 9
US-08-290-665A-135
Sequence 135, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

;; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
;; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
;; NUMBER OF SEQUENCES: 263
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: MORGAN & FINNEGAN
;; STREET: 345 PARK AVENUE
;; CITY: NEW YORK
;; STATE: NEW YORK
;; COUNTRY: USA
;; ZIP: 10154
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY DISK
;; COMPUTER: IBM PC COMPATIBLE
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WORDPERFECT 5.1
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/290,665A
;; FILING DATE: 15-AUG-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: RICHARD W. BORK
;; REGISTRATION NUMBER: 36,459
;; REFERENCE/DOCKET NUMBER: 2026-4116
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;; TELEX: 421792
;;
;; INFORMATION FOR SEQ. ID NO: 135:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 573 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;;
;; ORIGINAL SOURCE:
;; ORGANISM: homosapiens
;; INDIVIDUAL ISOLATE: HK10
;;
US-08-290-665A-135

Query Match 74.3%; Score 256.2; DB 2; Length 573;
Best Local Similarity 85.8%; Pred. No. 3e-62;

Matches 296; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 1 ATGAGCAGACTCTCTTAACCAAAAGAAACCAAAAGAAACCAACACC-CCGCCACAG 59
DB 1 ATGAGCAGACTCTCTTAACCAAAAGAAACCAAAAGAAACCAACACC-CCGCCACAG 60
QY 60 GACGTTAAGTCCAGAGCGGCGGCGAGATCGTGTGGAGTTTACGCTACACGCGAG 119
DB 61 GACGTTAAGTCCAGAGCGGCGGCGAGATCGTGTGGAGTTTACGCTACACGCGAG 120
QY 120 GGGCCCAAGTGGGTGCTGCTGACGTCAGCAAGACTTCCGAGCGGTGCAACCTCGCAGT 179
DB 121 GGGCCCAAGTGGGTGCTGCTGACGTCAGCAAGACTTCCGAGCGGTGCAACCTCGCAGT 180
QY 180 AGGCGCAACCCATCCAGAGGCGCGCGAAGCCGAGGCGAGTCTGGCTGACGCCGAG 239
DB 181 CGACGACAGCTATCCCAAGGCGCGCTCGAGCGAAGGCGCGTCTGAGCGCTCAGCCGAG 240
QY 240 TACCTTGGCCCTATATGGAATGAGGCTCGGCGTGGGCGAGGTGGCTCTGCTGCTCCG 299
DB 241 TACCTTGGCCCTATATGGAATGAGGCTCGGCGTGGGCGAGGTGGCTCTGCTGCTCCG 300
QY 300 CGGCGCTCTGCGCGTGGTGGGCGCAAAATGACCCCGCGCAGG 344
DB 301 CGGCGCTCTGCGCGTGGTGGGCGCAAAATGACCCCGCGCAGG 345

RESULT 10
US-08-290-665A-137
; Sequence 137, Application us/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BORK, J., MILLER, R.H. AND

;; APPLICANT: PURCELL, R.H.
;; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
;; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
;; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
;; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
;; NUMBER OF SEQUENCES: 263
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: MORGAN & FINNEGAN
;; STREET: 345 PARK AVENUE
;; CITY: NEW YORK
;; STATE: NEW YORK
;; COUNTRY: USA
;; ZIP: 10154
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY DISK
;; COMPUTER: IBM PC COMPATIBLE
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WORDPERFECT 5.1
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/290,665A
;; FILING DATE: 15-AUG-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: RICHARD W. BORK
;; REGISTRATION NUMBER: 36,459
;; REFERENCE/DOCKET NUMBER: 2026-4116
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;; TELEX: 421792
;;
;; INFORMATION FOR SEQ. ID NO: 137:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 573 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;;
;; ORIGINAL SOURCE:
;; ORGANISM: homosapiens
;; INDIVIDUAL ISOLATE: S2
;;
US-08-290-665A-137

Query Match 74.3%; Score 256.2; DB 2; Length 573;
Best Local Similarity 85.8%; Pred. No. 3e-62;

Matches 296; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 1 ATGAGCAGACTCTCTTAACCAAAAGAAACCAAAAGAAACCAACACC-CCGCCACAG 59
DB 1 ATGAGCAGACTCTCTTAACCAAAAGAAACCAAAAGAAACCAACACC-CCGCCACAG 60
QY 60 GACGTTAAGTCCAGAGCGGCGGCGAGATCGTGTGGAGTTTACGCTACACGCGAG 119
DB 61 GACATCAAGTCCCGGGGTGGCGGAGACATCGTGTGGAGTATACGTTGCGCGCAGG 120
QY 120 GGGCCCAAGTGGGTGCTGCTGACGTCAGCAAGACTTCCGAGCGGTGCAACCTCGCAGT 179
DB 121 GGGCCCAAGTGGGTGCTGCTGACGTCAGCAAGACTTCCGAGCGGTGCAACCTCGCAGT 180
QY 180 AGGCGCAACCCATCCAGAGGCGCGCGAAGCCGAGGCGAGTCTGGCTGACGCCGAG 239
DB 181 CGGCGACAGCTATCCCAAGGCGCGCTCGAGCGAAGGCGCGTCTGAGCGCTCAGCCGAG 240
QY 240 TACCTTGGCCCTATATGGAATGAGGCTCGGCGTGGGCGAGGTGGCTCTGCTGCTCCG 299
DB 241 TACCTTGGCCCTATATGGAATGAGGCTCGGCGTGGGCGAGGTGGCTCTGCTGCTCCG 300
QY 300 CGGCGCTCTGCGCGTGGTGGGCGCAAAATGACCCCGCGCAGG 344
DB 301 CGGCGCTCTGCGCGTGGTGGGCGCAAAATGACCCCGCGCAGG 345

RESULT 11
US-08-290-665A-138

```

Sequence 138, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK12
US-08-290-665A-138

```

```

Query Match      74.3%; Score 256.2; DB 2; Length 573;
Best Local Similarity 85.8%; Pred. No. 3e-62;
Matches 296; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
QY 1 ATGAGCACACTTCTTAACCAAGAAAACCAAAAGAAACCAACCAACG-CGGCCACAG 59
Db 1 ATGAGCACACTTCTTAACCAAGAAAACCAAAAGAAACCAACCAACG 60
QY 60 GACGTTAAGTTCAGAGCGGCGGCGGAGATGTTGGAGATTACGTGCTACACGACAG 119
Db 61 GACGTTAAGTTCAGAGCGGCGGCGGAGATGTTGGAGATTACGTGCTGCGCCAGG 120
QY 120 GGGCCAGATTGGGTGGCGTGCAGTCCGAAACTTCGAGCGGTGCGAACCTGCGACT 179
Db 121 GGGCCAGATTGGGTGGCGTGCAGTCCGAAACTTCGAGCGGTGCGAACCTGCGCGA 180
QY 180 AGCGCCCAACCCATCCCAAGGCGCGCAACGAGGAGAGGCTGCGGCGTCCAGCCGG 239
Db 181 CGGGGACAGCTATCCCAAGGCGCGCTCGGAGGAGGCGGTCTGAGCTGCGCGA 240
QY 240 TACCTTGGCCCTATATGAGATGAGGCGTGGGCTGAGGCTGCTGTCCTGCTCCCG 299
Db 241 TACCTTGGCCCTCTATGAGTACGAGGCTGGGCTGGGAGGCTGCTGCTGCTCCCA 300
QY 300 CGCGGCTCTGCGCGCTGCTGAGGCGCAAAATGAGCCCGCGCGCAGG 344
Db 301 CGCGGCTCTGCGCGCTGCTGAGGCGCAAAATGAGCCCGCGCGAGG 345

```

```

RESULT 12
PCT-US95-10398-135
Sequence 135, Application PC/US9510398
GENERAL INFORMATION:
APPLICANT: BUKH, J., MILLER, R.H. AND
PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 135:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: HK10
PCT-US95-10398-135

```

```

Query Match      74.3%; Score 256.2; DB 5; Length 573;
Best Local Similarity 85.8%; Pred. No. 3e-62;
Matches 296; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
QY 1 ATGAGCACACTTCTTAACCAAGAAAACCAAAAGAAACCAACCAACG-CGGCCACAG 59
Db 1 ATGAGCACACTTCTTAACCTCAAGAAAACCAAAAGAAACCAACCAACG 60
QY 60 GACGTTAAGTTCAGAGCGGCGGCGGAGATGTTGGAGATTACGTGCTACACGACAG 119
Db 61 GACGTTAAGTTCAGAGCGGCGGCGGAGATGTTGGAGATTACGTGCTGCGCCAGG 120
QY 120 GGGCCAGATTGGGTGGCGTGCAGTCCGAAACTTCGAGCGGTGCGAACCTGCGACT 179
Db 121 GGGCCAGATTGGGTGGCGTGCAGTCCGAAACTTCGAGCGGTGCGAACCTGCGCGA 180
QY 180 AGCGCCCAACCCATCCCAAGGCGCGCAAGAGGAGGCTGCTGAGCTGAGCCCGG 239

```

Db 181 CGACGACACCTATATGGAATGAGGCGGCGGAGCGAGCCGCTCTGGGCTCAGCCCCGG 240
QY 240 TACCCCTGGCCCCCTATATGGAATGAGGCGGCGGAGGCGGCTCTCTGCTCCCCG 299
Db 241 TACCCCTGGCCCCCTATATGGAATGAGGCGGCGGAGGCGGCTCTCTGCTCCCCA 300
QY 300 CGCGGCTCTGCGCCGCTGCTGGGCGCCCAATGACCCCCCGGCGCAGG 344
Db 301 CGCGGCTCTGCGCCGCTGCTGGGCGCCCAATGACCCCCCGGCGCAGG 345

RESULT 13
PCT-US95-10398-137
; Sequence 137, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: S2
; PCT-US95-10398-137

Query Match 74.3%; Score 256.2; DB 5; Length 573;
Best Local Similarity 85.8%; Pred. No. 3e-62;
Matches 296; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 1 ATAGACGACCTCTCTATATGGAATGAGGCGGCGGAGCGGCTCTGGGCTCAGCCCCG 59
Db 1 ATAGACGACCTCTCTATATGGAATGAGGCGGCGGAGCGGCTCTGGGCTCAGCCCCG 60
QY 60 GACGTTAAGTTCCAGCGCGGCTGAGATGCTTGGAATTACGCTTACACGCGCAGG 119

Db 61 GACATCAAGATTCCCGGGGCGGAGACAGATGCTTGAGATATACGTGTTGCCGCGCAGG 120
QY 120 GACCCCGCGGTTGGGAGTGTGTCAGTGCAGGCAAGACTTCCAGCGGTCGCACTCCAGT 179
Db 121 GACCCCGCGGTTGGGAGTGTGTCAGTGCAGGCAAGACTTCTGAAAGCTCAGGCTCCGGA 180
QY 180 AGCGGCGCAACCCATCCCAAGGCGCGCGCAACCGAAGGCAAGTCTGAGGCTCAGCGCGG 239
Db 181 CGCGGACACCTATATGGAATGAGGCGGCGGAGCGGAGCGGAGGCTCTGAGCCCGG 240
QY 240 TACCCCTGGCCCCCTATATGGAATGAGGCGGCGGAGGCGGCTCTCTGCTCCCCG 299
Db 241 TACCCCTGGCCCCCTATATGGAATGAGGCGGCGGAGGCGGCTCTCTGCTCCCCA 300
QY 300 CGCGGCTCTGCGCCGCTGCTGGGCGCCCAATGACCCCCCGGCGCAGG 344
Db 301 CGCGGCTCTGCGCCGCTGCTGGGCGCCCAATGACCCCCCGGCGCAGG 345

RESULT 14
PCT-US95-10398-138
; Sequence 138, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: DK12
; PCT-US95-10398-138


```

Query Match      74.3%; Score 256.2; DB 5; Length 573;
Best Local Similarity 85.8%; Pred. No. 3e-62;
Matches 296; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

1 ATGAGCAGACTTCCTTAACCAAGAAAAACCAAGAACCAACCAACCC-CGGCCACAG 59
1 ATGAGCAGACTTCCTTAACCAAGAAAAACCAAGAACCAACCAACCCAG 60
60 GACGTTAAGTTCAGAGCGCGGTGATGCTGTTGATGCTTACCTACACGACAG 119
61 GACGTTAAGTTCAGAGCGCGGTGATGCTGTTGATGCTTACCTACACGACAG 120
120 GCGCCAGTTCGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGT 179
121 GCGCCAGTTCGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGT 180
180 AGCGCCAGTTCGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGT 239
181 GCGCCAGTTCGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGT 240
240 TACCTTTGCGCTTATATGAGGATGAGGCTGCGGCTGCGGCTGCTGCTGCTG 299
241 TACCTTTGCGCTTATATGAGGATGAGGCTGCGGCTGCGGCTGCTGCTGCTG 300
300 GCGGCTCTGCGCTTATATGAGGATGAGGCTGCGGCTGCGGCTGCTGCTGCTG 344
301 GCGGCTCTGCGCTTATATGAGGATGAGGCTGCGGCTGCGGCTGCTGCTGCTG 345

RESULT 15
US-08-462-195-1
; Sequence 1, Application US/08462195
; Patent No. 5789544
; GENERAL INFORMATION:
; APPLICANT: MIYAMURA, TATSUO
; APPLICANT: SAITO, IZUMU
; APPLICANT: MATSUDERA, YOSHIHARU
; APPLICANT: HONDA, YOSHIKAZU
; APPLICANT: SEKI, MAKOTO
; TITLE OF INVENTION: METHOD FOR PRODUCING ECTOPROTEIN OF
; TITLE OF INVENTION: HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OHION, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESSER: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,195
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,303
; FILING DATE: 22-MAY-1995
; APPLICATION NUMBER: US 08/074,584
; FILING DATE: 11-JUN-1993
; APPLICATION NUMBER: JP 152487/1992
; FILING DATE: 11-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ohion, No. 5789544man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4169-003-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220

```

```

TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1037 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis C virus
; IMMEDIATE SOURCE:
; CLONE: pUC010
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 17..1036
; US-08-462-195-1

Query Match      74.3%; Score 256.2; DB 1; Length 1037;
Best Local Similarity 85.8%; Pred. No. 3.e-62;
Matches 296; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

1 ATGAGCAGACTTCCTTAACCAAGAAAAACCAAGAACCAACCAACCC-CGGCCACAG 59
17 ATGAGCAGCAAAATCCTTAACCAAGAAAAACCAAGAACCAACCAACCCAG 76
60 GACGTTAAGTTCAGAGCGCGGTGATGCTGTTGATGCTTACCTACACGACAG 119
77 GACGTTAAGTTCAGAGCGCGGTGATGCTGTTGATGCTTACCTACACGACAG 136
120 GCGCCAGTTCGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGT 179
137 GCGCCAGTTCGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGT 196
180 AGCGCCAGTTCGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGT 239
197 AGCGCAACCAATCCTTAACCAAGAAAAACCAAGAACCAACCAACCCAG 256
240 TACCTTTGCGCTTATATGAGGATGAGGCTGCGGCTGCGGCTGCTGCTGCTG 299
257 TACCTTTGCGCTTATATGAGGATGAGGCTGCGGCTGCGGCTGCTGCTGCTG 316
300 GCGGCTCTGCGCTTATATGAGGATGAGGCTGCGGCTGCGGCTGCTGCTGCTG 344
317 GCGGCTCTGCGCTTATATGAGGATGAGGCTGCGGCTGCGGCTGCTGCTGCTG 361

```

Search completed: February 19, 2003, 01:16:14
Job time : 81 secs

Thu Feb 20 14:41:31 2003

us-09-873-224-147.npb

Page 1

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 00:28:11 : Search time 105 Seconds
(without alignments)
1673.492 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 345
Sequence: 1 ATGAGCAGACTCTCTTAACCAACCCCGGCGCAGCA 345

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 424239 seqs, 254661826 residues

Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	334	96.8	346	9	US-09-899-046-147
2	309	89.6	309	9	US-09-851-138-49
3	261.6	75.8	652	9	US-09-851-138-59
4	260.2	75.4	499	9	US-09-899-046-165
5	254.6	73.5	499	9	US-09-899-046-163
6	253.6	73.5	498	9	US-09-899-046-163
7	253	73.3	2433	9	US-09-973-025-49
8	248.8	72.1	957	9	US-09-851-138-11
9	246.6	71.5	9416	9	US-10-104-966-13
10	246.6	71.5	9416	10	US-09-929-955-13
11	246.6	71.5	9646	10	US-09-995-937-1
12	246.6	71.5	9646	10	US-09-742-659-3
13	246.6	71.5	9646	10	US-09-238-076-1
14	246.6	71.5	10803	9	US-09-747-419-17
15	246.6	71.5	12980	9	US-09-995-937-5
16	246.6	71.5	12980	10	US-09-238-076-5
17	245	71.0	9379	9	US-09-916-359-1
18	245	71.0	9416	10	US-09-995-937-19
19	245	71.0	9416	10	US-09-238-076-19

20	243.4	70.6	630	10	US-09-973-322-1	Sequence 1, Appli
21	242.2	70.2	339	10	US-09-921-397-115	Sequence 115, App
22	241.8	70.1	831	9	US-09-851-138-65	Sequence 65, Appl
23	241.8	70.1	957	9	US-09-899-046-143	Sequence 143, App
24	238.6	69.2	957	9	US-09-851-138-43	Sequence 43, Appl
25	238.2	69.0	957	9	US-09-851-138-47	Sequence 47, Appl
26	235.4	68.2	957	9	US-09-851-138-17	Sequence 17, Appl
27	235	68.1	413	9	US-09-851-138-45	Sequence 45, Appl
28	233.4	67.7	499	9	US-09-899-046-151	Sequence 151, App
29	233.4	67.7	509	9	US-09-899-046-43	Sequence 43, Appl
30	233.4	67.7	509	9	US-09-899-046-43	Sequence 43, Appl
31	233.4	67.7	959	9	US-09-899-046-51	Sequence 51, Appl
32	233.4	67.7	959	9	US-09-899-046-51	Sequence 51, Appl
33	233.4	67.7	959	9	US-09-899-046-53	Sequence 53, Appl
34	230.8	66.9	327	9	US-09-851-138-1	Sequence 1, Appli
35	222	64.6	957	9	US-09-851-138-45	Sequence 45, Appl
36	222.6	64.5	1195	9	US-09-851-138-75	Sequence 75, Appl
37	222.6	64.5	310	9	US-09-851-138-21	Sequence 21, Appl
38	216	59.0	310	9	US-09-851-138-21	Sequence 21, Appl
39	203.6	58.9	310	9	US-09-851-138-13	Sequence 13, Appl
40	203.2	58.2	289	9	US-09-899-046-191	Sequence 191, App
41	200.8	58.2	310	10	US-09-921-397-114	Sequence 114, App
42	194.6	56.4	310	10	US-09-921-397-41	Sequence 41, Appl
43	169	49.0	234	10	US-09-851-138-9	Sequence 9, Appli
44	167.8	48.6	223	9	US-09-851-138-9	Sequence 9, Appli
45	137.4	39.8	461	9	US-09-851-138-103	Sequence 103, App

ALIGNMENTS

RESULT 1
US-09-899-046-147 Application US/09899046
; Sequence 147, Application US/09899046
; Publication No. US2003008274A1
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus
genotypes for diagnosis, prophylaxis and therapy.
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,046
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 147:
SEQUENCE CHARACTERISTICS:
LENGTH: 346 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEetical: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..346
NAME/KEY: matpeptide
LOCATION: 1..342
US-09-899-046-147
Query Match 96.8%; Score 334; DB 9; Length 346;
Best Local Similarity 99.7%; Pred. No. 4.5e-85;
Matches 345; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Oy 1 ATGAGCAGACTCTCTTAACCAACCAAGAAAGAACCAACCAACCCGCGCCACAG 59

Db 1 ATAGGACACCTTCTTAACACCAAGAAACCAAGAAACCAACCCCGGCGACAG 60
QY 60 GAGTTAAGTTCCAGGCGGCGTCAATCTTGTGAGATTACGTCTACCAACGAG 119
Db 61 GAGTTAAGTTCCAGGCGGCGTCAATCTTGTGAGATTACGTCTACCAACGAG 120
QY 120 GCGCCCAAGTTGGGTGCGTCAATCTTGTGAGATTACGTCTACCAACGAG 179
Db 121 GCGCCCAAGTTGGGTGCGTCAATCTTGTGAGATTACGTCTACCAACGAG 180
QY 180 AGAGCCCAACCCATCCCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 239
Db 181 AGAGCCCAACCCATCCCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
QY 240 TACCTTGGCCCTATATGGAATGAGGCTGCGGAGTGGCGAGGCTGCTCTCCCG 299
Db 241 TACCTTGGCCCTATATGGAATGAGGCTGCGGAGTGGCGAGGCTGCTCTCCCG 300
QY 300 GCGGCTCTCGCCCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 345
Db 301 GCGGCTCTCGCCCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 346

RESULT 2
US-09-851-138-49

; Sequence 49, Application US/09851138
; Publication No. US20020183508A1
GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
STUYVER, LIEVEN

TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
AGENTS

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE

STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075

FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9

FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7

FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:

NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 49:

SEQUENCE CHARACTERISTICS:
LENGTH: 309 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

MOLECULE TYPE: linear
TOPOLOGY: linear

HYPOHETICAL: NO
AMTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 49:
US-09-851-138-49

Query Match 89.6%; Score 309; DB 9; Length 309;
Best Local Similarity 100.0%; Pred. No. 4.7e-78;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATAGACACCTTCTTAACACCAAGAAACCAAGAAACCAACCCCGGCGACAG 60
Db 1 ATAGACACCTTCTTAACACCAAGAAACCAAGAAACCAACCCCGGCGACAG 60
QY 61 AGTTAAGTTCCAGGCGGCGTCAATCTTGTGAGATTACGTCTACCAACGAG 120
Db 61 AGTTAAGTTCCAGGCGGCGTCAATCTTGTGAGATTACGTCTACCAACGAG 120
QY 121 GCGCCCAAGTTGGGTGCGTCAATCTTGTGAGATTACGTCTACCAACGAG 180
Db 121 GCGCCCAAGTTGGGTGCGTCAATCTTGTGAGATTACGTCTACCAACGAG 180
QY 181 AGAGCCCAACCCATCCCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 181 AGAGCCCAACCCATCCCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
QY 241 ACCCTTGGCCCTATATGGAATGAGGCTGCGGAGTGGCGAGGCTGCTCTCCCG 300
Db 241 ACCCTTGGCCCTATATGGAATGAGGCTGCGGAGTGGCGAGGCTGCTCTCCCG 300
QY 301 GCGGCTCTC 309
Db 301 GCGGCTCTC 309

RESULT 3
US-09-851-138-59

; Sequence 59, Application US/09851138
; Publication No. US20020183508A1
GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
STUYVER, LIEVEN

TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
AGENTS

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE

STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075

FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9

FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7

FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:

NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775

REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:
LENGTH: 652 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

MOLECULE TYPE: linear
TOPOLOGY: linear

QY 120 GGCCTCCAGTTGGTGTCGTGACATGCGCAAGACTTCCAGCGGTGCAACCTCGCAGT 1
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GGCCCTAGAGTTGGTGGTGTGGCGCACACTCGAAGACTTGGAGCGGTGCGCAACCTCGTGGG 1

[illegible]

Db 61 GACGTCACTTCCGGGCGGTGTCAGATCGTTGTGGAGTTTACTCTTGGCCGGCAGG 120
 QY 120 GGGCCCCAGTTGGGTGTGCTGTCAGTGGCGCAAGACTTCCGAGGGTCCCAACCTCGCAGT 179
 Db 121 GGGCCCCAGTTGGGTGTGCTGTCAGTGGCGCAAGACTTCCGAGGGTCCCAACCTCGCAGG 180
 QY 180 AGCGCCCAACCATCCCGCAGGGCGCGCGCAACCGAGGGCAGGTCTCTGGGCTCAGCCCGGG 239
 Db 181 AGCGCCCAACCATCCCGCAGGGCGCGCGCAACCGAGGGCAGGTCTCTGGGCTCAGCCCGGG 240
 QY 240 TACCCCTTGGCCCTATATGGAATGAGGGCTGCGGTGGGCGAGGTGCTCTGTCGCCG 299
 Db 241 TACCCCTTGGCCCTATATGGAATGAGGGCTGCGGTGGGCGAGGTGCTCTGTCGCCG 300
 QY 300 CGGGGCTCTGGCCGTGTGGGGCCCAATGACCCCGGCGCAGG 344
 Db 301 CGGGGCTCTGGCCGTGTGGGGCCCAATGACCCCGGCGCAGG 345

RESULT 8

US-09-851-138-11
 ; Sequence 11, Application US/09851138
 ; Publication No. US20020183508A1

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
 STUDYER, LIEVEN

TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
 AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC

AGENTS

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:
 ADDRESSEE: ARNOLD, WHITE & DURKEE
 STREET: P.O. BOX 4433

CITY: HOUSTON
 STATE: TEXAS

COUNTRY: USA
 ZIP: 77210-4433

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word 6.0 / ASCII text output

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/851,138

FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/836,075

FILING DATE: <Unknown>
 APPLICATION NUMBER: EP 94870166.9

FILING DATE: 21 Oct 1994
 APPLICATION NUMBER: EP 95870076.7

FILING DATE: 28 Jun 1995

ATTORNEY/AGENT INFORMATION:
 NAME: KAMMERER, PATRICIA A.

REGISTRATION NUMBER: 29,775
 REFERENCE/DOCKET NUMBER: INNS:004

INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:

LENGTH: 957 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 11:

US-09-851-138-11

Query Match 72.1%; Score 248.8; DB 9; Length 957;
 Best Local Similarity 84.1%; Pred. No. 5.8e-61;

Matches 291; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 1 ATGAGCACTTCTTAACCAAGAAAAAACAAGAAACCAACCAAC-CCGGCCACAG 59

Db 1 ATGAGCACTTCTTAACCAAGAAAAAACAAGAAACCAACCAAC-CCGGCCACAG 60
 QY 60 GACGTTAAGTTCCAGCGCGCGGTGATGCTGTTGATGATTTACGCTACACAGCAG 119
 Db 61 GACGTTAAGTTCCAGCGCGCGGTGATGCTGTTGATGATTTACGCTACACAGCAG 120
 QY 120 GGGCCCCAGTTGGGTGTGCTGTCAGTGGCGCAAGACTTCCGAGGGTCCCAACCTCGCAGT 179
 Db 121 GGGCCCCAGTTGGGTGTGCTGTCAGTGGCGCAAGACTTCCGAGGGTCCCAACCTCGCAGG 180
 QY 180 AGCGCCCAACCATCCCGCAGGGCGCGCGCAACCGAGGGCAGGTCTCTGGGCTCAGCCCGGG 239
 Db 181 AGCGCCCAACCATCCCGCAGGGCGCGCGCAACCGAGGGCAGGTCTCTGGGCTCAGCCCGGG 240
 QY 240 TACCCCTTGGCCCTATATGGAATGAGGGCTGCGGTGGGCGAGGTGCTCTGTCGCCG 299
 Db 241 TACCCCTTGGCCCTATATGGAATGAGGGCTGCGGTGGGCGAGGTGCTCTGTCGCCG 300
 QY 300 CGGGGCTCTGGCCGTGTGGGGCCCAATGACCCCGGCGCAGG 343
 Db 301 CGGGGCTCTGGCCGTGTGGGGCCCAATGACCCCGGCGCAGG 344

RESULT 9

US-10-104-966-13
 ; Sequence 13, Application US/10104966
 ; Patent No. US2002015124A1

GENERAL INFORMATION:

APPLICANT: Matti Salberg

APPLICANT: Catharina Hultgren
 TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND

FILE REFERENCE: TRIPEP.23AUSC1

CURRENT APPLICATION NUMBER: US/10/104,966

CURRENT FILING DATE: 2002-03-22 547

PRIOR FILING DATE: 2000-11-03

PRIOR APPLICATION NUMBER: 60/229,175

PRIOR FILING DATE: 2000-08-29

NUMBER OF SEQ ID NOS: 15

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13

LENGTH: 9416

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Hepatitis C virus sequence

US-10-104-966-13

Query Match 71.5%; Score 246.6; DB 9; Length 9416;
 Best Local Similarity 84.1%; Pred. No. 4.9e-60;

Matches 290; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

QY 1 ATGAGCACTTCTTAACCAAGAAAAAACAAGAAACCAACCAAC-CCGGCCACAG 59

Db 342 ATGAGCACTTCTTAACCAAGAAAAAACAAGAAACCAACCAAC-CCGGCCACAG 401

QY 60 GACGTTAAGTTCCAGCGCGCGGTGATGCTGTTGATGATTTACGCTACACAGCAG 119

Db 402 GACGTTAAGTTCCAGCGCGCGGTGATGCTGTTGATGATTTACGCTACACAGCAG 461

QY 120 GGGCCCCAGTTGGGTGTGCTGTCAGTGGCGCAAGACTTCCGAGGGTCCCAACCTCGCAGT 179

Db 462 GGGCCCTAGATTGGGTGTGCTGTCAGTGGCGCAAGACTTCCGAGGGTCCCAACCTCGCAGT 521

QY 180 AGCGCCCAACCATCCCGCAGGGCGCGCGCAACCGAGGGCAGGTCTCTGGGCTCAGCCCGGG 239

Db 522 AGACGTACGCTATCCCAAGGACGCTCGGCCGAGGAGGACCTGGGCTCAGCCCGGG 581

QY 240 TACCCCTTGGCCCTATATGGAATGAGGGCTGCGGTGGGCGAGGTGCTCTGTCGCCG 299

Db 582 TACCCCTTGGCCCTATATGGAATGAGGGCTGCGGTGGGCGAGGTGCTCTGTCGCCG 641

OY 300 CGCGGCTCTGCGCCGCTGCGGGGCCCAATGACCCCGCGCAGG 344
 DB 642 CGTGGCTCTGCGGCTAGTGGGGCCCAACAGACCCCGCGCTAGG 686

RESULT 10

US-09-929-955-13
 ; Sequence 13, Application US/09929955
 ; Patent No. US20020136740A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Matti Salberg
 ; APPLICANT: Catharina Hultgren
 ; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
 ; FILE REFERENCE: TRIPEP-23AUS2
 ; CURRENT APPLICATION NUMBER: US/09/929,955
 ; PRIORITY FILING DATE: 2001-08-15
 ; PRIOR APPLICATION NUMBER: 09/705,547
 ; PRIOR FILING DATE: 2000-11-03
 ; PRIOR APPLICATION NUMBER: 60/229,175
 ; PRIOR FILING DATE: 2000-08-29
 ; PRIOR APPLICATION NUMBER: 60/225,767
 ; PRIOR FILING DATE: 2000-08-17
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 9416
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Hepatitis C virus sequence
 US-09-929-955-13

Query Match 71.5%; Score 246.6; DB 10; Length 9416;
 Best Local Similarity 84.1%; Pred. No. 4,9e-60;
 Matches 290; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

OY 1 ATGAGCACACTTCTTAACCAAGAAAAACAAAGAAACCAACCAACCC-CGCGCAGAG 59
 DB 342 ATGAGCACAGATCTTAACCTCAAGAAAAACCAAGCAACCAACCGCGCCAGAG 401
 OY 60 GACGTAAGTCTCCAGCGGGCGGCGCAGATCTTGTGAGTTTACGTCTACACGAGG 119
 DB 402 GACGTAAGTCTCCAGCGGGCGGCGCAGATCTTGTGAGTTTACGTCTACACGAGG 461
 OY 120 GGGCCCAAGTGGGTGTGCGTGCAGTGCAGCAAGACTTCCGAGCGGTGCGCAACCTGCGAGT 179
 DB 462 GGGCCCAAGTGGGTGTGCGTGCAGTGCAGCAAGACTTCCGAGCGGTGCGCAACCTGCGAGT 521
 OY 180 AGCGGCAACCAACCAACCAAGGCGCGCGCAACCGAGGCGAGGCTGAGCTCAGCCCGGG 239
 DB 522 AGAGGTGACCTATATCCCAAGGCGCGTGGCGCGGAGGAGGAGCTGGGCTCAGCCCGGG 581
 OY 240 TACCTTGGGCGCCCTATATGGAATGAGGCGGTGGGCGGAGGCTGCTCTGCTCCCG 239
 DB 582 TACCTTGGGCGCCCTATATGGAATGAGGCGGTGGGCGGAGGCTGCTCTGCTCCCG 239
 OY 300 CGCGGCTCTGCGCCGCTGCGGGGCCCAATGACCCCGCGCAGG 344
 DB 642 CGTGGCTCTGCGGCTAGTGGGGCCCAACAGACCCCGCGCTAGG 686

RESULT 11
 US-09-995-937-1
 ; Sequence 1, Application US/0995937
 ; Publication No. US20030028010A1
 ; GENERAL INFORMATION:
 ; APPLICANT: RICE, CHARLES et al.
 ; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
 ; VIRUS (HCV) AND USES THEREOF
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: HOWELL & HAERKAMP, L.C.
 STREET: 7733 FORSYTH BLVD., SUITE 1400
 CITY: ST. LOUIS
 STATE: MO
 COUNTRY: USA
 ZIP: 63105

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/995,937
 FILING DATE: 28-Nov-02
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/034,756
 FILING DATE: 04-May-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 REFERENCE/DOCKET NUMBER: 6029-4831
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 314-727-5188
 TELEFAX: 314-727-6092

SEQUENCE CHARACTERISTICS:
 INFORMATION FOR SEQ ID NO: 1:
 LENGTH: 9646 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO

US-09-995-937-1
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Query Match 71.5%; Score 246.6; DB 9; Length 9646;
 Best Local Similarity 84.1%; Pred. No. 5e-60;
 Matches 290; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

OY 1 ATGAGCACACTTCTTAACCAAGAAAAACCAAGAAACCAACCAACCC-CGCGCAGAG 59
 DB 342 ATGAGCACAGATCTTAACCTCAAGAAAAACCAAGCAACCAACCGCGCCAGAG 401
 OY 60 GACGTAAGTCTCCAGCGGGCGGCGCAGATCTTGTGAGTTTACGTCTACACGAGG 119
 DB 402 GACGTAAGTCTCCAGCGGGCGGCGCAGATCTTGTGAGTTTACGTCTACACGAGG 461
 OY 120 GGGCCCAAGTGGGTGTGCGTGCAGTGCAGCAAGACTTCCGAGCGGTGCGCAACCTGCGAGT 179
 DB 462 GGGCCCAAGTGGGTGTGCGTGCAGTGCAGCAAGACTTCCGAGCGGTGCGCAACCTGCGAGT 521
 OY 180 AGCGGCAACCAACCAACCAAGGCGCGCGCAACCGAGGCGAGGCTGAGCTCAGCCCGGG 239
 DB 522 AGAGGTGACCTATATCCCAAGGCGCGTGGCGCGGAGGAGGAGCTGGGCTCAGCCCGGG 581
 OY 240 TACCTTGGGCGCCCTATATGGAATGAGGCGGTGGGCGGAGGCTGCTCTGCTCCCG 239
 DB 582 TACCTTGGGCGCCCTATATGGAATGAGGCGGTGGGCGGAGGCTGCTCTGCTCCCG 239
 OY 300 CGCGGCTCTGCGCCGCTGCGGGGCCCAATGACCCCGCGCAGG 344
 DB 642 CGTGGCTCTGCGGCTAGTGGGGCCCAACAGACCCCGCGCTAGG 686

RESULT 12
 US-09-742-659-3
 ; Sequence 3, Application US/09742659
 ; Patent No. US20010034019A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hong, Zhi
 ; APPLICANT: Butkiewicz, Nancy J.

APPLICANT: Zhong, Weidong
 APPLICANT: Ingravallo, Paul
 APPLICANT: Wright-Minogue, Jacquelyn
 APPLICANT: Lau, Johnson Y.
 APPLICANT: Lemon, Stanley M.
 TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
 FILE REFERENCE: ID01116
 CURRENT APPLICATION NUMBER: US/09/742,659
 PRIOR FILING DATE: 2000-12-21
 PRIOR APPLICATION NUMBER: US 60/171,469
 NUMBER OF SEQ ID NOS: 16
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 3
 LENGTH: 9646
 TYPE: DNA
 ORGANISM: Hepatitis C virus
 US-09-742-659-3

Query Match 71.5% Score 246.6; DB 10; Length 9646;
 Best Local Similarity 84.1%; Pred. No. 5e-60;
 Matches 290; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

QY 1 ATGAGCACTTCTTAACCAAGAAAAAGAAAAAGAACCAACCC-CGGCCACAG 59
 Db 342 ATGAGCACTTCTTAACCTCAAGAAAAAGAAAAAGAACCAACCCGCGCCACAG 401
 QY 60 GACCTTAAGTTCCAGCGCGCGGTCAGATCGTTGGTGAAGTTACGTACACGACAG 119
 Db 402 GACCTCAAGTTCCGCGGCGCGGTCAGATCGTTGGTGAAGTTACGTACGCGCAGG 461
 QY 120 GACCCCAAGTTGGTGTGCGTGCAGTGCAGACCTCCAGCGGTCGCAACCTCGAGT 179
 Db 462 GCGCCCTAGATTGGTGTGCGCGCGGTCAGAGAGAACTTCGAGCGGTCGCAACCTCGAGT 521
 QY 180 AGCGCCCAACCTCCAGCGCGCGCGGTCAGAGAGAACTTCGAGCGGTCGCAACCTCGAGT 239
 Db 522 AGAGCTGAGCTATCCCAAGGACGTCGCGCGGTCAGAGAGAACTTCGAGCGGTCGCAACCTCGAGT 581
 QY 240 TACCTTGGCCCTATATGGAATGAGGGTGGGGTGGGTCAGAGGTCGCTGTCCTCCG 239
 Db 582 TACCTTGGCCCTCTATATGGAATGAGGGTGGGGTGGGTCAGAGGTCGCTGTCCTCCG 641
 QY 300 CGCGGCTCTGCGCCGTCGTGGGGCCCAATGACCCCGCGCGCAG 344
 Db 642 CGTGGCTCTGCGCTAGCTGGGGCCCAAGACCCCGCGCGTAG 686

RESULT 13
 US-09-238-076-1
 ; Sequence 1, Application US/09238076
 ; Patent No. US20020102540A1
 ; GENERAL INFORMATION:
 ; APPLICANT: RICE, CHARLES et al.
 ; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
 ; STREET: 7733 FORSYTH BLVD., SUITE 1400
 ; CITY: ST. LOUIS
 ; STATE: MO
 ; COUNTRY: USA
 ; ZIP: 63105
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/238,076
 ; FILING DATE:
 ; CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/034,756
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: HOLLAND, DONALD R.
 REGISTRATION NUMBER: 35,197
 REFERENCE/DOCKET NUMBER: 6029-4831
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 314-727-5188
 TELEFAX: 314-727-6092
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9646 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-09-238-076-1

Query Match 71.5% Score 246.6; DB 10; Length 9646;
 Best Local Similarity 84.1%; Pred. No. 5e-60;
 Matches 290; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

QY 1 ATGAGCACTTCTTAACCAAGAAAAAGAAAAAGAACCAACCC-CGGCCACAG 59
 Db 342 ATGAGCACTTCTTAACCTCAAGAAAAAGAAAAAGAACCAACCCGCGCCACAG 401
 QY 60 GACCTTAAGTTCCAGCGCGCGGTCAGATCGTTGGTGAAGTTACGTACACGACAG 119
 Db 402 GACCTCAAGTTCCGCGGCGCGGTCAGATCGTTGGTGAAGTTACGTACGCGCAGG 461
 QY 120 GACCCCAAGTTGGTGTGCGTGCAGTGCAGACCTCCAGCGGTCGCAACCTCGAGT 179
 Db 462 GCGCCCTAGATTGGTGTGCGCGCGGTCAGAGAGAACTTCGAGCGGTCGCAACCTCGAGT 521
 QY 180 AGCGCCCAACCTCCAGCGCGCGCGGTCAGAGAGAACTTCGAGCGGTCGCAACCTCGAGT 239
 Db 522 AGAGCTGAGCTATCCCAAGGACGTCGCGCGGTCAGAGAGAACTTCGAGCGGTCGCAACCTCGAGT 581
 QY 240 TACCTTGGCCCTATATGGAATGAGGGTGGGGTGGGTCAGAGGTCGCTGTCCTCCG 239
 Db 582 TACCTTGGCCCTCTATATGGAATGAGGGTGGGGTGGGTCAGAGGTCGCTGTCCTCCG 641
 QY 300 CGCGGCTCTGCGCCGTCGTGGGGCCCAATGACCCCGCGCGCAG 344
 Db 642 CGTGGCTCTGCGCTAGCTGGGGCCCAAGACCCCGCGCGTAG 686

RESULT 14
 US-09-747-419-17
 ; Sequence 17, Application US/09747419
 ; Patent No. US2002015582A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lemon, Stanley
 ; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
 ; FILE REFERENCE: 265,007 0101
 ; CURRENT APPLICATION NUMBER: US/09/747,419
 ; PRIOR FILING DATE: 2000-12-23
 ; PRIOR APPLICATION NUMBER: US 60/171,909
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 17
 ; LENGTH: 10803
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Nucleotide sequence of MK0-2
 ; US-09-747-419-17

Query Match	71.5%	Score 246.6	DB 9	Length 10803
Best Local Similarity	84.1%	Pred. No. 5.1e-60		
Matches 290	Conservative	0	Mismatches 54	Indels 1
				Gaps 1
QY 1	ATGACACACTCTCCCTAAACACACAAGAAAAACCAAAAGAACCCACACC-C-GGGCCACAG 59			
Db 342	ATGACACAGCAATCCTAAACCTCAAGAAAAAACCAAAAGCAATACACCAACCGTGGCCACAG 401			
QY 60	GACGTTAAGTCCACAGGCGCGGTGAGATCGTTGTGAGTTTACGTGCTACACCCAG 119			
Db 402	GACGTCAAGTTCGCCGGGTGGCGGTGAGATCGTTGTGAGTTTACTGTTGGCCGGCAG 461			
QY 120	GGCCCCCAGTTGGGTGTGGTGCACGTGGCGACAGACTTCCGACCGGTCCGCAACTGCGAGT 179			
Db 462	GGCCCTAGATTGGGTGTGGTGGCGGACGACAGAAAGACTTCCGACCGGTCCGCAACTGAGGT 521			
QY 180	AGGCGCCAAACCCATCCCGACAGGGCGGCCGCAACCGAGGCGAGTCTCGGCTCAGGCCCGGG 239			
Db 522	AGAGCTCAAGCCTATCCCGCAAGGACAGTGGGGCCGAGGGCAGGACTCGGGCTCAGCCCGGG 581			
QY 240	TACCCCTTGGCCCTTATATGGCAATAGAGGCTCGGGGTGGGCAAGGGTGGCTCTGTGCCG 299			
Db 582	TACCCCTTGGCCCTTATATGGCAATAGAGGCTTCGGGTGGGCGGGATAGCTCTGTCTCC 641			
QY 300	CGCGGCTCGCGCCGTCGNGGGGCCCAATAGACCCCGGCGAG 344			
Db 642	CGTGCGCTCGGCGCTAGCTGGGGGCCCCACAGACCCCGCGGTAGG 686			

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

: SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-995-37-5

Query Match 71.5%; Score 246.6; DB 9; Length 12980;
Best Local Similarity 84.1%; Pred. No. 5,46-60;
Matches 290; Conservative 0; Mismatches 54; Indels 1; Gaps 1.

OY 1 ATAGGCACTTCTCTTAACACCAAGAAAAAACCAAAACACCAACCC-CGGCCACAG 59
|||||
Db 342 ATGAGCAAGATCTTAACCTCAAAAGAAAAACCAAACTTAACACCAACCGTGGCCAC 401
60 GACGTTAAGTCCAGAGCGCGGTGATCGTTGTGGAGTTTACGATACCAACGAGG 119
|||||
Db 402 GACGTCAGTCTCCCGGGGTGCGGTGATCGTGTGGTGAAGTTTACTTGTTCGGCGAGG 461
120 GGCCCCAGTTGGGTGTGCGTGCAGTGCAGAGACTTCGAGCGGTGCGCAACTGCGAGT 179
|||||
Db 462 GGCCCTAGATTGGGTGTGGCGCGGAGAGAGAGACTTCGAGGCGGTGCGCAACTGAGGT 521
180 AGGCGCCAAACCCATCCCAAGGCGCGCGAGACCGAGAGGCGAGTCTGTGGCTCAACCCGGG 239
522 AGACGTAGAGCTATGCCCAAGGACACTCGCCCGAGGCGAGAGACTTGGGTCTCAACCCGGG 581
240 TACCGTTGGCCCTATATGGGAATGAGGCGTGGGTGGGCGAGGCTGGTCTGTCCCG 299
582 TACCGTTGGCCCTATATGGCAATGAGGCTTCCGGGTGGGCGGAGTGGCTCTCTCTCC 641
300 CGCGGCTCGCCGTCGTGGTGGGCGCCAAATGACCCCGGCGCAGG 344
642 CGTGGCTCTCGGCTACTGTGGGCGCCCAAGACCCCGGCGTAGG 686

```

Search completed: February 19, 2003, 01:18:18
Job time : 115 secs

```

: SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-995-37-5

Query Match 71.5%; Score 246.6; DB 9; Length 12980;
Best Local Similarity 84.1%; Pred. No. 5,46-60;
Matches 290; Conservative 0; Mismatches 54; Indels 1; Gaps 1.

OY 1 ATAGGCACTTCTCTTAACACCAAGAAAAAACCAAAACACCAACCC-CGGCCACAG 59
|||||
Db 342 ATGAGCAAGATCTTAACCTCAAAAGAAAAACCAAACTTAACACCAACCGTGGCCAC 401
60 GACGTTAAGTCCAGAGCGCGGTGATCGTTGTGGAGTTTACGATACCAACGAGG 119
|||||
Db 402 GACGTCAGTCTCCCGGGGTGCGGTGATCGTGTGGTGAAGTTTACTTGTTCGGCGAGG 461
120 GGCCCCAGTTGGGTGTGCGTGCAGTGCAGAGACTTCGAGCGGTGCGCAACTGCGAGT 179
|||||
Db 462 GGCCCTAGATTGGGTGTGGCGCGGAGAGAGAGACTTCGAGGCGGTGCGCAACTGAGGT 521
180 AGGCGCCAAACCCATCCCAAGGCGCGCGAGACCGAGAGGCGAGTCTGTGGCTCAACCCGGG 239
522 AGACGTAGAGCTATGCCCAAGGACACTCGCCCGAGGCGAGAGACTTGGGTCTCAACCCGGG 581
240 TACCGTTGGCCCTATATGGGAATGAGGCGTGGGTGGGCGAGGCTGGTCTGTCCCG 299
582 TACCGTTGGCCCTATATGGCAATGAGGCTTCCGGGTGGGCGGAGTGGCTCTCTCTCC 641
300 CGCGGCTCGCCGTCGTGGTGGGCGCCAAATGACCCCGGCGCAGG 344
642 CGTGGCTCTCGGCTACTGTGGGCGCCCAAGACCCCGGCGTAGG 686

```



```

PE 27-APR-1994: 94WO-EP01323.
XX
XX 27-APR-1993: 93EP-0401099.
PR 05-AUG-1993: 93EP-0402019.
XX
PA (INNO-) INNOGENETICS NV SA.
XX
PI Maertens G, Stuyver L;
XX
DR WPI: 1994-358277/44.
XX N-PSDB: AAQ78089.
XX
PT New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates
XX
XX Claim 11: Page 198-199; 404pp; English.
PS
CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the core region
CC of HCV.
XX
XX Sequence 115 AA:
SQ
Alignment Scores:
Pred. No.: 3,62e-54 Length: 115
Score: 617.00 Matches: 115
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 96.55% Indels: 0
DB: Gaps: 0
US-09-873-224-147 (1-345) x AAR63351 (1-115)
OY 1 ATGAGCAGCACTTCTTAACCAAGAAAACCAAAACCAACCCGCGCACAG 60
DB 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrAsnProGlyHisArg 20
OY 61 ACGTTAAGTTCCAGCGCGGCGTGCAGATCGTGTGAGATTACGTACACGACGAG 120
DB 21 ThrLeuSerSerGlnAlaAlaValArgSerLeuValGlnPheThrCysTyrHisAlaGly 40
OY 121 GCCCCCAAGTGGGTGGTGCATGCGCAGACTTCGAGCGGTGCGCAACCTCGCACTA 180
DB 41 AlaProSerTrpValCysValGlnCysAlaArgLeuProSerGlyArgAsnLeuAlaVal 60
OY 181 GGGGCAACCCATCCCGCAGGCGCGCCGACGAGGAGGAGGCTCGGCGTACCCGGGT 240
DB 61 GlyAlaAsnProSerProGlyArgAlaGlnProArgAlaGlyProGlyLeuSerProGly 80
OY 241 ACCCTTGGCCCTATATGGGAATGAGGCGTGGGCGGAGGAGGCTCGTCCGTCGCCGC 300
DB 81 ThrLeuGlyProTyrMetGlyMetArgAlaAlaGlyGlyGlnGlySerCysProArg 100
OY 301 GCGGCTCTGCGCCGCTGTGGGGCCCAATGACCCCGCGGAGGA 345
DB 101 AlaAlaLeuAlaArgArgGlyAlaGlnMetThrProGlyAlaGly 115
RESULT 2
AAR6550
ID AAR6550 standard: peptide: 115 AA.

```

```

XX AC AAR6550;
XX AC
XX 11-MAR-1997 (first entry)
XX
XX Hepatitis C virus type 10a isolate NE98 amino acids 1-317.
XX
XX Hepatitis C virus; subtype: polymerase chain reaction; amplification;
XX PCR; primer: probe; antibody; infection.
XX
XX Hepatitis C virus.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 17
XX FT /note="amino acid residue at this position is not known"
XX
XX W09613590-A2.
XX
XX 09-MAY-1996.
XX
XX 23-OCT-1995: 95WO-EP04155.
XX
XX 28-JUN-1995: 95EP-0870076.
XX 21-OCT-1994: 94EP-0870166.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Maertens G, Stuyver L;
XX
XX WPI: 1996-251460/25.
XX N-PSDB: AAT27961.
XX
XX Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
XX - used to develop probes and primers for new sub:types and vaccines
XX to prevent and treat infection
XX
XX Claim 25; Fig 3; 150pp; English.
XX
XX The sequences AAR96526-R96578 represent novel sequences isolated from
XX hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
XX 4a-j, 5a and 6a, they esp. from the novel subtypes 1d-f, 2e-l, 2k, 2l,
XX 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
XX untranslated region (UR), the Core/E1, NS4 or NS5B regions of the
XX genome. This sequence represents amino acids 1-115 from the HCV type 10a
XX isolate NE98.
XX
XX The new HCV types were isolated from patients with chronic HCV from the
XX Benelux countries, France, Cameroon and Vietnam, because of their
XX aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
XX amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
XX sequenced either directly or partially and used to classify the new
XX viruses into (sub)types based on comparison with known sequences.
XX The sequences were used to generate the peptides AAR96424-R96524. The
XX sequences can also be used to synthesise probes and primers for the
XX detection of HCV in a sample. The polypeptides can be used to detect
XX anti-HCV antibodies, for HCV typing or to prevent HCV infections.
XX
XX Sequence 115 AA:
SQ
Alignment Scores:
Pred. No.: 2.95e-53 Length: 115
Score: 608.00 Matches: 114
Percent Similarity: 99.13% Conservative: 0
Best Local Similarity: 99.13% Mismatches: 1
Query Match: 95.15% Indels: 1
DB: Gaps: 0
US-09-873-224-147 (1-345) x AAR96550 (1-115)
OY 1 ATGAGCAGCACTTCTTAACCAAGAAAACCAAAACCAACCC-CGCGCACAG 59
DB 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrAsn***ArgProGln 20
OY 60 GACGTTAAGTTCCAGCGCGGCGTGCAGATCGTGTGAGATTACGTCTACACGACAG 119

```


Alignment Scores:

Pred. No.: 3 42e-51 Length: 191
 Score: 588.00 Matches: 108
 Percent Similarity: 96.52% Conservative: 3
 Best Local Similarity: 93.91% Mismatches: 4
 Query Match: 92.02% Indels: 1
 DB: 17 Gaps: 0

US-09-873-224-147 (1-345) x AAR92969 (1-191)

QY 1 ATGAGCACACTTCTTAACCAAGAAAGAAACCAACCA-CCCCGGCCACAG 59
 |||||||
 Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgSerThrLysArgProGln 20
 60 GACGTTAGTTCACAGGGGGGGGCGGTCAGATCGTGTGAGTTTACGTACACAGCAGG 119
 |||||||
 Db 21 AspValLysPheProGlyGlyGlnIleValIleGlyValIleValLeuProArg 40
 QY 120 GGCCCCAGTTGGGTGTCGTGAGTGGCGCAAGACTTCGAGCGGTCGCAACTCGCAGT 179
 |||||||
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
 QY 180 AGCGCCCAACCCACCCCGGCGGCGGCAAGCGAGGTCCTCGGCGTCAGCCGCGG 239
 |||||||
 Db 61 ArgArgGlnProLeuProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80
 QY 240 TACCTTGCCCTATATGGAATGAGGCGTGCAGGTGGGAGGCGGTCTCTGTCGCCG 299
 |||||||
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
 QY 300 CGCGGCTCTCGCCGTCGTGCGGCGCCCAATGACCCCGGCGCAGG 344
 |||||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 5

AAR92971

ID AAR92971 standard; Protein: 191 AA.

XX AAR92971;

XX 02-OCT-1996 (first entry)

XX Hepatitis C virus isolate DK12 core protein.

XX HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX Hepatitis.

XX Hepatitis C virus.

XX WO9605315-A2.

XX 22-FEB-1996.

XX 15-AUG-1995; 95WO-US10398.

XX 15-AUG-1994; 94US-0290665.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX (USSH) US SEC DEPT HEALTH.

XX Bukh J, Miller RH, Purcell RH;

XX WPI: 1996-139709/14.

XX N-PSDB; AAT16645.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins -

XX used to determine HCV genotype and as vaccines against HCV infection

XX Claim 4: Page 209-210; 340pp; English.

XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV

XX isolates. Isolated cDNA sequences are used for the prodn. of primers

CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may
 CC be used in the prevention of HCV infection.

SQ Sequence 191 AA:

Alignment Scores:
 Pred. No.: 3 42e-51 Length: 191
 Score: 588.00 Matches: 108
 Percent Similarity: 96.52% Conservative: 3
 Best Local Similarity: 93.91% Mismatches: 4
 Query Match: 92.02% Indels: 1
 DB: 17 Gaps: 0

US-09-873-224-147 (1-345) x AAR92971 (1-191)

QY 1 ATGAGCACACTTCTTAACCAAGAAAGAAACCAACCA-CCCCGGCCACAG 59
 |||||||
 Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgSerThrLysArgProGln 20
 60 GACGTTAGTTCACAGGGGGGGGCGGTCAGATCGTGTGAGTTTACGTACACAGCAGG 119
 |||||||
 Db 21 AspValLysPheProGlyGlyGlnIleValIleGlyValIleValLeuProArg 40
 QY 120 GGCCCCAGTTGGGTGTCGTGAGTGGCGCAAGACTTCGAGCGGTCGCAACTCGCAGT 179
 |||||||
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
 QY 180 AGCGCCCAACCCACCCCGGCGGCGGCAAGCGAGGTCCTCGGCGTCAGCCGCGG 239
 |||||||
 Db 61 ArgArgGlnProLeuProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80
 QY 240 TACCTTGCCCTATATGGAATGAGGCGTGCAGGTGGGAGGCGGTCTCTGTCGCCG 299
 |||||||
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
 QY 300 CGCGGCTCTCGCCGTCGTGCGGCGCCCAATGACCCCGGCGCAGG 344
 |||||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 6

AAR92970

ID AAR92970 standard; Protein: 191 AA.

XX AAR92970;

XX 02-OCT-1996 (first entry)

XX Hepatitis C virus isolate S2 core protein.

XX HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX Hepatitis.

XX Hepatitis C virus.

XX WO9605315-A2.

XX 22-FEB-1996.

XX 15-AUG-1995; 95WO-US10398.

XX 15-AUG-1994; 94US-0290665.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX (USSH) US SEC DEPT HEALTH.

XX Bukh J, Miller RH, Purcell RH;

XX WPI: 1996-139709/14.

XX N-PSDB; AAT16644.

Db	1812	GI ProArguenuGlyValAlaArgAlaThrArgIysThrSerGluArgSerGlnProArgGly	1831
Qy	180	AGCGCCCAACCCATCCCCAGGGCGCCGACCCAGGGCAGCTCTGGCTGCACGCCGGG	239
Db	1832	ArgArgGlnProIleProLysAlaArgAArgProGluGlyAqIlnrHPAlaGlnProGly	1851
Oy	240	TACCCTTGGCCCCCTATATGCGAATAGAGGCGCGGGGCGAGGGTGGCTCTGTCGCCG	299
Db	1852	TyrProItrProLeuItrLysnGlnGlyCysGlyTTPAlaGlyTTPLeuItrSerPro	1871
Qy	300	CGCGGCTCTCGCCCGTCGTGGGGCCCAATGACCCCGCGCCAGG	344
Db	1872	ArgGlySerArgProSerTrpGlyProThrAspProArgArgArg	1886
RESULT 10			
AAB62638	1D	AAB62638 standard; Protein; 1911 AA.	
XX	XX	AAB62638:	
XX	XX	23-JUL-2001 (first entry)	
XX	XX	Amino acid sequence of pd.deltansNS5.pj.core140.	
XX	XX	HCV; mutant; non-structural protein; NS; hepatitis C virus; mutation;	
XX	XX	catalytic domain; NS3; NS4; NS5; antiviral; vaccine; immunostimulant;	
XX	XX	immunotherapy.	
XX	XX	Synthetic.	
XX	XX	Hepatitis C virus.	
XX	XX	MO200138360-A2.	
XX	XX	31-MAY-2001.	
XX	XX	22-NOV-2000; 2000MO-US32326.	
XX	XX	24-NOV-1999; 99US-0167502.	
XX	XX	(CHIR) CHIRON CORP.	
XX	XX	Colt D, Medina-Selby A, Selby M, Houghton M;	
XX	XX	WPI: 2001-343948/36.	
XX	XX	N-PSDB: AAF83674.	
XX	XX	Mutant non-structural (NS) Hepatitis C virus (HCV) polypeptide, useful	
XX	XX	as a vaccine against HCV, comprises a polypeptide having a mutation	
XX	XX	that functionally disrupts the catalytic domain of NS3 -	
XX	XX	Example 1; Fig 21; 340pp; English.	
XX	XX	The invention relates to an isolated mutant non-structural (NS) Hepatitis	
XX	XX	C virus (HCV) polypeptide, comprising a polypeptide having a mutation in	
XX	XX	the catalytic domain of NS3, where the mutation functionally disrupts the	
XX	XX	catalytic domain. The NS mutant polypeptides can include NS3, NS4 (NS4a	
XX	XX	and NS4b) NS5 (NS5a and NS5b) or portions thereof. The HCV polypeptide	
XX	XX	and polynucleotide (preferably DNA or a plasmid) compositions can be used	
XX	XX	in vaccines against HCV and as diagnostics. The antibodies raised against	
XX	XX	CC in these polypeptides can also be used as diagnostics, or for passive	
XX	XX	immunotherapy. The antibodies are also useful for isolating and	
XX	XX	identifying HCV particles. The present sequence represents the amino	
XX	XX	acid sequence of the pd.deltansNS5.pj.core140.	
XX	XX	Sequence 1911 AA:	
Alignment Scores:			
Pred. No.:	1,07e-49	Length:	1911
Score:	575.00	Matches:	105
Percent Similarity:	94.78%	Conservative:	4
Best Local Similarity:	91.30%	Mismatches:	6
Query Match:	89.98%	Indels:	1

```

DB:                22                Gaps:                0
US-09-873-224-147 (1-345) x AAB62638 (1-1911)
OY      1      ATGAGCACACTTCCATAACACCAAGAAAAACCAAGAAACACCAACC-CGGCCACAG 59
         |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1772  MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGln 1791
OY      60      GACGTTAACTTCCACAGCGCGCGGTCCAGATCCGTTGGTGAGTTTACGTGTACCAACGACAG 119
         |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1792  AspValLysPheProGlyGlyGlyGlnIleValIGlyGlyValTyrLeuLeuProArgArg 1811
OY      120      GGGCCCCACTTGGGTGGTGGTGGTCAGNGCGCAGACACTTCCAGACGGTCCGACCTCGCACT 179
         |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1812  GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 1831
OY      180      AGCGCCCAACCCATCCCGCAGGGCGCGCCGACACCGAGGCGAGTCTCCTGGGCTACGCCCGG 239
         |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1832  ArgArgGlnProIleProLysAlaArgArgProGluGlyArgThrProAlaGlnProGly 1851
OY      240      TACCTTGGCCCTATATGGAATGAGGGCTCGGGTGGGCGAGGGGCGTCCGTCGCCG 299
         |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1852  TyrProThrProLeuTyrGlyAsnGlyCysGlyTyrAlaGlyTyrLeuLeuSerPro 1871
OY      300      CGGGGCTCTGCCCGCCGTCGGGGCCCAATGACCCCGCCGCGAGG 344
         |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1872  ArgGlySerArgProSerTyrGlyProThrAspProArgArgArg 1886

RESULT 11
AAB62639
ID      AAB62639 standard; Protein; 1921 AA.
AC      AAB62639;
XX
XX      23-JUL-2001 (first entry)
DT
XX
XX      Amino acid sequence of pd.deltans3NS5.pj.core150.
DE
KW      HCV; mutant; non-structural protein; NS; hepatitis C virus; mutation;
KM      catalytic domain; NS3; NS4; NS5; antiviral; vaccine; immunostimulant;
KW      immunotherapy.
OS      Synthetic.
OS      Hepatitis C virus.
PN      WC200138360-A2.
XX
XX      31-MAY-2001.
PD
XX
XX      22-NOV-2000; 2000MO-US32326.
PF
XX
XX      24-NOV-1999; 9905-0167502.
PR
XX
XX      (CHIR ) CHIRON CORP.
PA
XX
XX      Coit D, Medina-Selby A, Selby M, Houghton M;
PI
XX
XX      WPI; 2001-343948/36.
DR      N-PSDB; AAF83675.
XX
XX      Mutant non-structural (NS) Hepatitis C virus (HCV) polypeptide, useful
PT      as a vaccine against HCV, comprises a polypeptide having a mutation
PT      that functionally disrupts the catalytic domain of NS3 -
XX
XX      Example 1; Fig 22; 340pp; English.
XX
XX      The invention relates to an isolated mutant non-structural (NS) Hepatitis
CC      C virus (HCV) polypeptide, comprising a polypeptide having a mutation in
CC      the catalytic domain of NS3, where the mutation functionally disrupts the
CC      catalytic domain. The NS mutant polypeptides can include NS3, NS4 (NS4a
CC      and NS4b) NS5 (NS5a and NS5b) or portions thereof. The HCV polypeptide
CC      and polynucleotide (preferably DNA or a plasmid) compositions can be used
CC      in vaccines against HCV and as diagnostics. The antibodies raised against

```


XX hepatitis.

OS Hepatitis C virus.

FN W09605315-A2.

PD 22-FEB-1996.

PF 15-AUG-1995; 95WC-USI0398.

PR 15-AUG-1994; 94US-0290665.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
(USSH) US SEC DEPT HEALTH.

P1 Bukh J, Miller RH, Purcell RH;

DR WPt: 1996-139709/14.

DR N-PSDB; AATL648.

XX DNA and amino acid sequence of HCV envelope 1 and core proteins -

PT used to determine HCV genotype and as vaccines against HCV infection

XX Claim 4; Page 212; 340pp; English.

PS
CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
CC isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC can also useful for HCV genotyping. Proteins encoded by the cDNAs
CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may
CC be used in the prevention of HCV infection.

SQ Sequence 191 AA;

Alignment Scores:
Pred. No.: 8,96e-50 Length: 191
Score: 574.00 Matches: 106
Percent Similarity: 95.65% Conservative: 4
Best local Similarity: 92.17% Mismatches: 5
Query Match: 89.83% Indels: 1
DB: Gaps: 0

US-09-873-224-147 (1-345) x AAR92974 (1-191)

OY 1 ATGAGCACACTCTTAACCAACAAGAATAAACCACAAGAACCAACCG-CGGCCACAG 59
||||| |||||
Db 1 MetSerThrAsnProLysProGIlnArglySthrLySarqsnTrhrAsnArGrPromet 20

OY 60 GACGTAAAGTTCGCCGGCGGCGGCACAGTCTCTTGAGATTACGNGCYNACCAAGG 119
|||||
Db 21 AspvAlvSpnePrroGLyglVgLnLeValdJlgyLVallYrFleulePProArGaig 40

OY 120 GCCCCCAGTTGGGTGTCGTGCAGTAGTGCGAAGACTTCGAGCGGTGCAACCTGCAGT 179
|||||
Db 41 GLYPROArGlueGLyValArGalALalArLyShrserGIunArqserGINProArGLy 60

OY 180 AGCGCCCAACCATCCCCAGAGCGCGCGCAAACCGAGAGGAGTCTTGCGCTCAGCCGGG 239
|||||
Db 61 ArgArGInPrroLIerProLYsaLAarArqrserGIunLyArqrserTPPALagINProGLy 80

OY 240 TACCGTTGACCCCTATATGGAGATAGAGGCTGCGSGTGGGAGAGGTGCCTCTGTCGCCG 299
|||||
Db 81 TyrPrOTrrProLeutrYcIlAsmclucILygselYrrPalaaGLYrrPlleuleuSerPro 1000

OY 300 CGCGGCTCTCGCCGTCGTGGGGCCCCAAATGACCCCGCGCGCAGG 344
|||||
Db 101 ArglySerArqProSerTrIpGLyProAsnsPraoArGrArqarg 115

RESULT 15
AAR92976
ID AAR92976 standard; Protein: 191 AA.

Alignment Scores:
 Pred. No.: 8,12e-49 Length: 114
 Score: 579.00 Matches: 106
 Percent Similarity: 96.49% Conservative: 4
 Best Local Similarity: 92.98% Mismatches: 4
 Query Match: 90.61% Indels: 1
 DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x S41361 (1-114)

QY 1 ATGAGCACACTTCTTAACCAAGAAAGAAACCAAGAACCAACCA-CCCGGGCCACAG 59
 |||||
 Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrLeuArgProGln 20
 QY 60 GACGTTAAGTTCACAGCGCGGCTCAGATCGTTGGTGAAGTTTACGTCTACACAGCGAG 119
 |||||
 Db 21 AspValLysPheProGlyGlyGlnIleValGlyValValLeuProArgArg 40
 QY 120 GCGCCCGAGTTGGTGGTGGCGAGTGGCGAGACTCCGAGCGGTCCGACCTCCAGT 179
 |||||
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
 QY 180 AGGCGCCAAACCCATCCAGAGGCGCGCGCAACGAGGCGAGTCTCGGCTCAGCCCGGG 239
 |||||
 Db 61 ArgArgGlnProIleProLysAlaArgGlnSerGlnIleValArgSerTrpPalaGlnProGly 80
 QY 240 TACCTTGGCCCCATATAGGGAATGAGGGCGGTGGCGAGGTGGCTCTGTCCTCCCG 299
 |||||
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyGlyGlyTrpPalaGlyTrpLeuLeuSerPro 100
 QY 300 CCGGCTCTCGCCCTCTGTCGGGCGCCCAATGACCCCGCGCGC 341
 |||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArg 114

RESULT 5

S41361 genome polypeptide - hepatitis C virus (genotype 3, N4) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 3, N4
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41361
 R:Van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 Submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.
 A:Reference number: S41341
 A:Accession: S41361
 A:Molecule type: genomic RNA
 A:Residues: 1-123 <N4>
 A:Cross-references: EMBL:Z29464; NID:q443890; PIDN:CA82602.1; PID:q443891
 A:Experimental source: genotype 3, N4
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; polypeptide
 F:1-123/Product: core protein #status predicted <MAY>

Alignment Scores:
 Pred. No.: 1.02e-48 Length: 123
 Score: 578.00 Matches: 105
 Percent Similarity: 95.65% Conservative: 5
 Best Local Similarity: 91.30% Mismatches: 5
 Query Match: 90.45% Indels: 1
 DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x S41361 (1-123)

QY 1 ATGAGCACACTTCTTAACCAAGAAAGAAACCAAGAACCAACCA-CCCGGGCCACAG 59
 |||||
 Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrLeuArgProGln 20
 QY 60 GACGTTAAGTTCACAGCGCGGCTCAGATCGTTGGTGAAGTTTACGTCTACACAGCGAG 119
 |||||
 Db 21 AspValArgPheProGlyGlyGlnIleValGlyValValLeuProArgArg 40

QY 120 GCGCCCGAGTTGGTGGTGGCGAGTGGCGAGACTCCGAGCGGTCCGACCTCGCAGT 179
 |||||
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
 QY 180 AGGCGCCAAACCCATCCAGAGGCGCGCGCAACGAGGCGAGTCTCGGCTCAGCCCGGG 239
 |||||
 Db 61 ArgArgGlnProIleProLysAlaArgGlnSerGlnIleValArgSerTrpPalaGlnProGly 80
 QY 240 TACCTTGGCCCCATATAGGGAATGAGGGCGGTGGCGAGGTGGCTCTGTCCTCCCG 299
 |||||
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyGlyGlyTrpPalaGlyTrpLeuLeuSerPro 100
 QY 300 CCGGCTCTCGCCCTCTGTCGGGCGCCCAATGACCCCGCGCGC 344
 |||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArg 115

RESULT 6

PC2061 genome polypeptide N2 - hepatitis C virus

N:Contains: envelope protein E1; nonstructural protein E2/NS1
 C:Species: hepatitis C virus
 C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000
 C:Accession: PC2061
 R:Li, J.S.; Valtitski, L.; Tong, S.P.; Trepo, C.
 Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
 A:Title: Identification of the third major genotype of hepatitis C virus in France.
 A:Reference number: PC2060; MUID:94197744; PMID:8147893
 A:Accession: PC2061
 A:Molecule type: mRNA
 A:Residues: 1-411 <L1J>
 A:Cross-references: GB:112355; NID:q410169; PIDN:AAA20155.1; PID:q410170
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein
 F:192-383/Product: envelope protein E1 #status predicted <SEP>
 F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NEP>
 F:384-209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
 Pred. No.: 1.23e-48 Length: 411
 Score: 577.00 Matches: 106
 Percent Similarity: 95.65% Conservative: 4
 Best Local Similarity: 92.17% Mismatches: 5
 Query Match: 90.30% Indels: 1
 DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x PC2061 (1-411)

QY 1 ATGAGCACACTTCTTAACCAAGAAAGAAACCAAGAACCAACCA-CCCGGGCCACAG 59
 |||||
 Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrLeuArgProGln 20
 QY 60 GACGTTAAGTTCACAGCGCGGCTCAGATCGTTGGTGAAGTTTACGTCTACACAGCGAG 119
 |||||
 Db 21 AspValLysPheProGlyGlyGlnIleValGlyValValLeuProArgArg 40
 QY 120 GCGCCCGAGTTGGTGGTGGCGAGTGGCGAGACTCCGAGCGGTCCGACCTCGCAGT 179
 |||||
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnTrpPalaGlnProArgGly 60
 QY 180 AGGCGCCAAACCCATCCAGAGGCGCGCGCAACGAGGCGAGTCTCGGCTCAGCCCGGG 239
 |||||
 Db 61 ArgArgGlnProIleProLysAlaArgGlnSerGlnIleValArgSerTrpPalaGlnProGly 80
 QY 240 TACCTTGGCCCCATATAGGGAATGAGGGCGGTGGCGAGGTGGCTCTGTCCTCCCG 299
 |||||
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyGlyGlyTrpPalaGlyTrpLeuLeuSerPro 100
 QY 300 CCGGCTCTCGCCCTCTGTCGGGCGCCCAATGACCCCGCGCGC 344
 |||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArg 115

RESULT 7

S41364

genome polypeptide - hepatitis C virus (genotype 4, N3) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 4, N3

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41364

R:Van Doorn, L.J.; Kleber, G.E.M.; Brouwer, J.T.

Submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.

A:Reference number: S41341

A:Accession: S41364

A:Molecule type: genomic RNA

A:Residues: 1-115 <V>

A:Cross-references: EMBL:Z29467; NID:9443896; PIDN:CAA82605.1; PID:9443897

A:Experimental source: genotype 4, N3

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: capsid protein; core protein; polypeptide

F:1-115/Product: core protein #status predicted <MAT>

Alignment Scores:

Pred. No.:	2,52e-48	Length:	115
Score:	574.00	Matches:	106
Percent Similarity:	95.65%	Conservative:	4
Best Local Similarity:	92.17%	Mismatches:	5
Query Match:	89.83%	Indels:	1
		Gaps:	0

US-09-873-224-147 (1-345) x S41364 (1-115)

QY 1 ATGACGACACTTCTTAACCAAGAAAGAAACCAACCAACC-CGGCCACAG 59

Db 1 MetserThrsnProLysProGlnArgLysThrLysArgAsnThrsnArgArgProMet 20

QY 60 GACGTTAAGTCCAGCGCGGCGTCAAGTCTGTGGATTACGCTACACAGCAGG 119

Db 21 AspValLysPheProGlyGlyGlnIleValIcylValIleuLeuProArgArg 40

QY 120 GGGCCCGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 179

Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60

QY 180 AGGCGCCACCAATCCCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 239

Db 61 ArgArgGlnProLeuProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80

QY 240 TACCCCTGGCCCTATATGGAATGAGGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGG 299

Db 81 TyrProTrpProLeuLysGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 100

QY 300 CGGCGCTCTCGCCGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 344

Db 101 ArgGlySerArgProSerTrpPolypProAsnAspProArgArgArg 115

RESULT 8

PC2060

genome polypeptide N1 - hepatitis C virus

N:Contains: envelope protein E1; nonstructural protein E2/NS1

C:Species: hepatitis C virus

C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 17-Nov-2000

C:Accession: PC2060

Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994

A:Title: Identification of the third major genotype of hepatitis C virus in France.

A:Reference number: PC2060; MUID:94197744; PMID:814783

A:Accession: PC2060

A:Molecule type: mRNA

A:Residues: 1-411 <LIP>

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;

F:192-383/Product: envelope protein E1 #status predicted <SPE>

F:384-411/Product: nonstructural protein E2/NS1 #status predicted <NDE>

F:196,209,234,305,325/Binding site: carboxylate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.:	3.82e-48	Length:	411
Score:	572.00	Matches:	104
Percent Similarity:	95.65%	Conservative:	6
Best Local Similarity:	90.43%	Mismatches:	5
Query Match:	89.51%	Indels:	1
		Gaps:	0

US-09-873-224-147 (1-345) x PC2060 (1-411)

QY 1 ATGACGACACTTCTTAACCAAGAAAGAAACCAACCAACC-CGGCCACAG 59

Db 1 MetserThrsnProLysProGlnArgLysThrLysArgAsnThrsnArgArgProMet 20

QY 60 GACGTTAAGTCCAGCGCGGCGTCAAGTCTGTGGATTACGCTACACAGCAGG 119

Db 21 AspValLysPheProGlyGlyGlnIleValIcylValIleuLeuProArgArg 40

QY 120 GGGCCCGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 179

Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60

QY 180 AGGCGCCACCAATCCCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 239

Db 61 ArgArgGlnProLeuProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80

QY 240 TACCCCTGGCCCTATATGGAATGAGGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGG 299

Db 81 TyrProTrpProLeuLysGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 100

QY 300 CGGCGCTCTCGCCGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 344

Db 101 ArgGlySerArgProSerTrpPolypProAsnAspProArgArgArg 115

RESULT 9

S41366

genome polypeptide - hepatitis C virus (genotype 4, N5) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 4, N5

C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

C:Accession: S41366

R:Van Doorn, L.J.; Kleber, G.E.M.; Brouwer, J.T.

Submitted to the EMBL Data Library, January 1994

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.

A:Reference number: S41341

A:Accession: S41366

A:Molecule type: genomic RNA

A:Residues: 1-114 <V>

A:Cross-references: EMBL:Z29469; NID:9443900; PIDN:CAA82607.1; PID:9443901

A:Experimental source: genotype 4, N5

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: capsid protein; core protein; polypeptide

F:1-114/Product: core protein #status predicted <MAT>

Alignment Scores:

Pred. No.:	7.8e-48	Length:	114
Score:	569.00	Matches:	105
Percent Similarity:	95.61%	Conservative:	4
Best Local Similarity:	92.11%	Mismatches:	5
Query Match:	89.05%	Indels:	1
		Gaps:	0

US-09-873-224-147 (1-345) x S41366 (1-114)

QY 1 ATGACGACACTTCTTAACCAAGAAAGAAACCAACCAACC-CGGCCACAG 59

Db 1 MetserThrsnProLysProGlnArgLysThrLysArgAsnThrsnArgArgProMet 20

QY 60 GACGTTAAGTCCAGCGCGGCGTCAAGTCTGTGGATTACGCTACACAGCAGG 119

Db 21 AspValLysPheProGlyGlyGlnIleValIcylValIleuLeuProArgArg 40

S41346

genome polyprotein - hepatitis C virus (genotype 1, N6) (fragment)

N:Contains: core protein

C:Species: hepatitis C virus

A:Variety: genotype 1, N6

C/Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000

A:Accession: S41346

R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.

A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LIPA.

A:Reference number: S41341

A:Accession: S41346

A:Molecule type: genomic RNA

A:Residues: 1-118 <YAN>

A:Cross-references: EMBL:Z29449; NID:9443860; PIDN:CAA82587.1; PID:9443861

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: capsid protein; core protein; polyprotein

F:1-118/Product: core protein #status predicted <MAN>

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-09-873-224-147 (1-345) x S41346 (1-118)	9.77e-48	118	366.00	105	4	6	0
Percent Similarity:	366.00		94.78%				
Best Local Similarity:	94.78%		91.30%				
Query Match:	88.89%						
DB:	2						

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-0							

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: February 19, 2003, 01:22:20 : Search time 21 Seconds

(without alignments)
1362.794 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 639
Sequence: 1 ATGACACACACTTCCCTAACC.....AATGACCCCGCGCAGGA 345

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+.n2p.model -DEV=slp
-Q=/cgn2.1/USPTO.spool/US09873224/runat_14022003_092704_27439/app.query.fasta.1.519
-DB=SwissProt.40 -OPMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USBR=US09873224.ecgn2.1.16.6runat_14022003_092704_27439 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568	88.9	3011	1	POLG_HCVH
2	560	87.6	3011	1	POLG_HCVI
3	558	87.3	513	1	POLG_HCVJ2
4	558	87.3	3010	1	POLG_HCVTA
5	558	87.3	3010	1	POLG_HCVJT
6	556	87.0	3010	1	POLG_HCVBK
7	550	86.1	520	1	POLG_HCVH4
8	550	86.1	520	1	POLG_HCVHK
9	542	84.8	737	1	POLG_HCVJ7
10	542	84.8	737	1	POLG_HCVJ8
11	541	84.7	737	1	POLG_HCVJ5
12	541	84.7	3033	1	POLG_HCVJ6
13	532	83.3	3010	1	POLG_HCVTW
14	101	15.8	750	1	ELS_CRICK
15	94	14.7	445	1	SM50_STRPU
16	92.5	14.5	407	1	SM41_HMPU
17	91.5	14.3	1147	1	MYSB_ACACA
18	91	14.2	188	1	GAG_AVEV2

19	91	14.2	453	1	GAG_AVIMD	P06444	avian myelo
20	90.5	14.2	900	1	IF2_MYCTU	P71613	mycobacteri
21	90	14.1	440	1	GAGC_AVISC	P05433	avian sarco
22	90	14.1	660	1	YHL1_EBY	P03181	epstein-bar
23	89	13.9	453	1	GAG_AVIMC	P03323	avian myelo
24	89	13.9	463	1	ANX7_MOUSE	007076	mus musculu
25	88.5	13.8	1733	1	VNUA_PRYKA	P33485	pseudorabi
26	88	13.8	786	1	PRGR_CHICK	P07812	gallus gall
27	88	13.7	1446	1	IE18_PRYKA	P33479	pseudorabi
28	87.5	13.7	240	1	RSMB_MONDO	09166	monodelphis
29	87	13.5	1083	1	T2D3_HUMAN	000268	homo sapien
30	86.5	13.5	289	1	CC19_CAEEL	P18835	Caenorhabd
31	86.5	13.5	591	1	FZD9_HUMAN	000144	homo sapien
32	86.5	13.5	688	1	BOMD_MOUSE	054839	mus musculu
33	86	13.5	309	1	GAG_FUJIV	P03326	fujinam sa
34	86	13.5	807	1	AFSK_STRGR	P54742	streptomyce
35	85.5	13.4	1323	1	NME4_MOUSE	003391	mus musculu
36	85	13.3	1168	1	MYSC_ACACA	P10509	acanthamoeb
37	85	13.3	1466	1	CA13_HUMAN	P02461	homo sapien
38	84	13.1	431	1	UL61_HCMVA	P16818	human cytom
39	84	13.1	686	1	BOMD_HUMAN	095936	homo sapien
40	84	13.1	924	1	IF2_MYCLE	P1675	pseudorabi
41	84	13.1	1461	1	IE18_PRYIF	09q30	rattus norv
42	84	13.1	2471	1	MYC2_RAT	062645	rattus norv
43	83.5	13.1	1323	1	NME4_RAT	P52875	mus musculu
44	83	12.9	323	1	PF27_MOUSE	096684	homo sapien
45	83	13.0	422	1	GF11_HUMAN		

ALIGNMENTS

RESULT 1

ID POLG_HCVH STANDARD: PRT: 3011 AA.

AC P27958: 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (core protein) (P22):
DE Envelope glycoprotein E1 (GP32) (GP35): Envelope glycoprotein E2
DE (GP68) (GP70) (NS1): Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-): Protease/helicase NS3 (P70) (Hepacivirus)
DE (EC 3.4.21.98): Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate H) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052256; PubMed=1658800;
RA Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA Prince A.M.;
RT "Genomic structure of the human prototype strain H of hepatitis C
RT virus: comparison with American and Japanese isolates."
RT Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
[2]
RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
DE MEDLINE=97331322; PubMed=9187654;
RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT "Structure of the hepatitis C virus RNA helicase domain."
RT Nat. Struct. Biol. 4:463-467(1997).
[3]
RX X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
DE MEDLINE=96154321; PubMed=9493270;
RA Kim J.L., Morgenstern K.A., Grifith J.P., Dwyer M.D., Thomson J.A.,
RA Murcko M.A., Lin C., Caron P.R.;
RT "Hepatitis C virus NS3 RNA helicase domain with a bound
RT oligonucleotide: the crystal structure provides insights into the mode
RT of unwinding."
RT Structure 6:89-100(1998).
CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.

CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position, Cys or Thr in P1 and Ser or Ala in P1',
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MNNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M67463; AAA45534.1; -
 DR PIR: A36814; GNMVCH.
 DR PDB: 1HEI; 25-NOV-98.
 DR PDB: 1AIV; 16-FEB-99.
 DR MEROPS: S29.001; -
 DR MEROPS: U39.001; -
 DR TRANSFAC: T04155; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; HCV_RdRp; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR ProDom: PD186062; HCV_NSI; 1.
 DR PolyProtein: glycoprotein; Transferrase; RNA-directed RNA polymerase;
 DR Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 DR Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 DR 3D-structure.
 KW INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 191
 FT CHAIN 192 383 CAPSID PROTEIN C.
 FT CHAIN 384 746 ENVELOPE GLYCOPROTEIN E1.
 FT CHAIN 747 809 ENVELOPE GLYCOPROTEIN E2.
 FT CHAIN 810 1026 PROTEIN P7.
 FT CHAIN 1027 1657 NONSTRUCTURAL PROTEIN NS2.
 FT CHAIN 1658 1711 PROTEASE/HELICASE NS3.
 FT CHAIN 1712 1972 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1972 1972 NONSTRUCTURAL PROTEIN NS4B.

FT CHAIN 1973 2420 NONSTRUCTURAL PROTEIN NS5A.
 FT CHAIN 2421 3011 NONSTRUCTURAL PROTEIN NS5B.
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
 Alignment Scores:
 Pred. No.: 6 38e-39 Length: 3011
 Score: 568.00 Matches: 105
 Percent Similarity: 94.78% Conservative: 4
 Best Local Similarity: 91.30% Mismatches: 6
 Query Match: 88.89% Indels: 1
 DB: Gaps: 0
 US-09-873-224-147 (1-345) x POLG_HCVH (1-3011)
 QY 1 ATGAGCACACTTCTTAACCAACAAAGAAACCAAGAAACCAACCAACC-CGGCCACAG 59
 Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgSntThrAsnArgArgProGln 20
 QY 60 GACGTTAAGTCCCGACGGCGCGGTCACATCGTTGGTGGAGTTTACGTACACACGAGG 119
 Db 21 AspValLysPheProGlyGlyGlyGlnIleValGlyValGlyLeuPheProArgArg 40
 QY 120 GGCCCCAGTTGGGGTGTGCTGCAGTCCGCAAGACTCCGACGGCGTGCACCTCGAGT 179
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerLysArgGlnProArgGly 60
 QY 180 AGGCGCCAACCCATCCCGAGGGCGCGGCGGAACGAGGGCGAGTCTGGGCTCAGCCGGG 239
 Db 61 ArgArgGlnProLysProLysAlaArgArgProGlnLysArgThrProArgGlnProGly 80
 QY 240 TACCCCTGGCCCCCATATGGAATGAGAGGCTGGGGTGGGCGAGGGTGCCTGTCCCGG 299
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTyrPheLeuTyrLeuLeuSerPro 100
 QY 300 CGCGCTCTCGCCCGCGTGGGGGCGGCAATGACCCCGCGGCGAGG 344
 Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArgArg 115
 RESULT 2
 POLG_HCV1 STANDARD; PRT; 3011 AA.
 AC P26664;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin)
 DE (EC 3.4.21.-); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP66) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus (isolate HC-12) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCB1_Taxid=11111;
 RN [1]
 RP MEDLINE-92230232; PubMed=1314459;
 RA Okamoto H., Kunita K., Okada S.T., Yamamoto K., Lizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishihiro S.;
 RT Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.*;
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D10074; BAA0968.1;
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC DR Polyprotein: Glycoprotein; Coat protein; Envelope protein;
 CC Transmembrane; Nonstructural protein.
 CC KW Transmembrane; Nonstructural protein.
 CC INTI_MET 1
 CC FT CHAIN 1
 CC FT CHAIN 115
 CC FT CHAIN 116
 CC FT CHAIN 191
 CC FT CHAIN 192
 CC FT CHAIN 383
 CC FT CHAIN 384
 CC FT CHAIN >513
 CC FT TRANSMEM 347
 CC FT TRANSMEM 369
 CC FT CARBOHYD 196
 CC FT CARBOHYD 209
 CC FT CARBOHYD 233
 CC FT CARBOHYD 234
 CC FT CARBOHYD 250
 CC FT CARBOHYD 250
 CC FT CARBOHYD 305
 CC FT CARBOHYD 417
 CC FT CARBOHYD 417
 CC FT CARBOHYD 423
 CC FT CARBOHYD 430
 CC FT CARBOHYD 448
 CC FT CARBOHYD 448
 CC FT NON_TER 513
 CC SO SEQUENCE 513 AA; 55704 MW; 943931E514CDEF3 CRC64;
 CC
 CC Alignment Scores:
 CC Pred. No.: 4,46e-38
 CC Score: 558.00
 CC Percent Similarity: 93.91%
 CC Best Local Similarity: 90.43%
 CC Query Match: 87.32%
 CC DB: 1
 CC Gaps: 0

QY 1 ATGAGCAGACTTCTCTAAACCCAGAAAGAAACCAAGAACCAACCC- CGGCCACAG 59
 DB 1 MetserThrasProLysProGlnArgLysThrLysArgSntnRasnArgArgProGln 20
 QY 60 GACGTTAGTCCAGGCGGCGGTACAGTCGTTGAGTTTACGTCACAGCGAG 119
 DB 21 AspValLysPheProGlyGlyGlyGlnIleValGlyValLysLeuLeuProArgArg 40
 QY 120 GGGCCCAAGTGGGTGTGCTGTCAGTGCAGACAGCTTCGAGCGTCCAGT 179
 DB 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
 QY 180 AGGGCCCAACCCAGCCAGCGCGCCGCAACGAGGAGGAGGCTGCTGAGCCGG 239
 DB 61 ArgArgGlnProLysProLysAlaArgArgProGlnIleValArgAlaThrArgLys 80
 QY 240 TACCTTGAGCCCTATATGGAATGAGGCTGCGGCGGAGGCTGCTGCTCCG 299
 DB 81 TyrProThrProLeuThrArgLysArgGlnGlyGlyGlyGlyGlyGlyGlyGlyGly 100
 QY 300 CGCGGCTCTGCGCCGCTGCGGCGCCCAATGACCCCGCGCAGG 344
 DB 101 ArgLysSerArgProSerThrGlyProThrArgProArgArgArg 115
 RESULT 4
 ID POLG_HCVJA STANDARD; PRT; 3010 AA.
 AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP66) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepcivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate Japanese) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCB1_Taxid=111116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91088550; PubMed=2175903;
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
 RA Sugimura T., Shimotohno K.;
 RT Molecular cloning of the human hepatitis C virus genome from
 RT Japanese patients with non-A, non-B hepatitis.*;
 RT Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE-91192160; PubMed=1849488;
 RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., MuraIso K.,
 RA Ohkoshi S., Shimotohno K.;
 RT Molecular structure of the Japanese hepatitis C viral genome.*;
 RT FEBS Lett. 280:325-328(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA] (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

FT	CARBOHYD	623	623	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	645	645	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	2041	2041	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	2077	2077	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	2240	2240	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	2529	2529	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	2788	2788	N-LINKED (GLCNAC . . .) (POTENTIAL)
SQ	SEQUENCE	3010 AA:	326573 MW:	94ALIC77435D64ZBB CRC64;

Alignment Scores:			
Pred. No.:	4,12e-38	Length:	3010
Score:	558.00	Matches:	104
Percent Similarity:	92.17%	Conservative:	2
Best Local Similarity:	90.43%	Mismatches:	9
Query Match:	87.32%	Indels:	1
DB:	1	Gaps:	0

US-09-873-224-147 (1-345) x POLG_HCVJUT (1-3010)			
QY	1	ATAGACACACTTCTTAACACAGAAMAACAAGAACACCACCA-CCCGGCCACAG	59
Dd	1	MeSerThrAsnProLysProGlnArgIsthrLysArgAsnThrTyArgArgProGln	20
QY	60	GAGCTTAAGTCCCAGCGCGCGGTGAGATCGTTGGTGAGTTACGTCTACACGACAG	119
Dd	21	AspValAlaShneProLeLyelgylIntlelValgllylValItyValleuProArgArg	40
QY	120	GGCCCCAATTGGGTGGTCGTGCACATGCCGAAGACTTCCGAGCGGTGCAACCTCGCAGT	179
Dd	41	GlYPTrHleuglYalArglarHlarHrArgIsthserGIuaRgserrInProArggly	60
QY	180	AGGCGGCACCATCCACAGGGCGCGCCGAACCGAGGCAAGCTTCGGCTCACGCCGG	239
Dd	61	ArgArgGlnProIleerPolysAlarArgArgProGluArgAlatPrIagInProgly	80
QY	240	TACCTTGGCCCTATATGGAGATAGGCGTCGGCGTGGGAGAGGTGCCTCGTCCCG	299
Dd	81	TyrProTrpProLeuTyrgLYasnGluGlyleuIltyrrPrIaaGlylrpleuLeuSerPro	100
QY	300	CGCGGTCTCGCCGCTCGTGGGCCCCAAATGACCCCGCGGCAG	344
Dd	101	ArgGlySerArgProSerTrpGlyProThrAspProArgArg	115

RESULT 6			
POLG_HCVBK			
ID	POLG_HCVBK	STANDARD:	PRT: 3010 AA.
AC	P26653:		
Dt	01-AUG-1992 (Rel. 23, Created)		
Dt	01-AUG-1992 (Rel. 23, Last sequence update)		
Dt	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Genome polyprotein [Contents: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (hepacivirin) (EC 3.4.21.99); Nonstructural protein NS4A (P4); Nonstructural protein NSB (P67); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].		
OS	Hepatitis C virus (isolate BK) (HCV).		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.		
OX	NCBI_TaxId=11105;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91140698; PubMed=1847440;		
RA	Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J., Oishi E., Andoh T., Yoshida I., Okayama H.; "Structure and organization of the hepatitis C virus genome isolated from human carriers"; J. Virol. 65:1105-1113(1991).		
RT	[2]		
RN	SEQUENCE OF 1487-1500.		
RX	MEDLINE=96251224; PubMed=8647104;		
RA	Borowski P., Heiland M., Oehlmann K., Becker B., Kornetovsk I.,		

RESULT 7			
ID	POLG_HCVH4	STANDARD:	PRT, 520 AA.
AC	POLG_HCVH4		
DT	01-JUL-1993 (Rel. 26, Created)		
DT	01-JUL-1993 (Rel. 26, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Genome polyprotein (Contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1)) (Fragment).		
DE	Hepatitis C virus (isolate HCV-4/76) (HCV).		
OC	viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.		
OX	NCBI_TaxID=31643;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93019030; PubMed=1383400;		
RA	Abe K., Inchauste G., Fujisawa K.;		
RT	"Genomic characterization and mutation rate of hepatitis C virus isolated from a patient who contracted hepatitis during an epidemic of non-A, non-B hepatitis in Japan."		
RL	J. Gen. Virol. 73:2725-2729(1992).		
CC	-1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).		
CC	-----		
DR	EMBL: D10688; BAA01530.1; -		
DR	InterPro: IPR002531; HCV_NSI.		
DR	InterPro: IPR002522; HCV_capsid.		
DR	InterPro: IPR002521; HCV_core.		
DR	InterPro: IPR002519; HCV_env.		
DR	Pfam: PF01539; HCV_env; 1.		
DR	Pfam: PF01542; HCV_core; 1.		
DR	Pfam: PF01543; HCV_capsid; 1.		
DR	Pfam: PF01560; HCV_NSI; 1.		
DR	ProDom: PD186062; HCV_NSI; 1.		
KW	Polyprotein; Glycoprotein; Coat protein; Envelope protein; Transmembrane; Nonstructural protein.		
FT	INIT_MET 1		
FT	CHAIN 1		
FT	CHAIN 1		
FT	CHAIN 116		
FT	CHAIN 191		
FT	CHAIN 383		
FT	CHAIN 192		
FT	CHAIN >520		
FT	TRANSMEM 347		
FT	CARBOHYD 196		
FT	CARBOHYD 209		
FT	CARBOHYD 234		
FT	CARBOHYD 305		
FT	CARBOHYD 418		
FT	CARBOHYD 424		
FT	CARBOHYD 431		
FT	CARBOHYD 449		
FT	NON_TER 520		
SO	SEQUENCE 520 AA; 56499 MW; AA135246CF20D525 CRC64;		
Alignment Scores:			
Pred. No.:	1.98e-37	Length:	520
Score:	550.00	Matches:	103
Percent Similarity:	93.04%	Conservative:	4
Best local Similarity:	89.57%	Mismatches:	8
Query Match:	86.07%	Indels:	1
DB:	1	Gaps:	0

```

US-09-873-224-147 (1-345) x POLG_HCVH4 (1-520)

QY      1 ATAGACACACTTCTCTTAACCAACAAAGAAAAACCAAGAACACCAACC-CGCCACAG  59
Db      1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln  20
QY      60 GAGCTTAAGTCTCCAGCGCGCGGTGAGTCGTTGGTGAGGATTACTGCTACCAACGAG  119
Db      21 AspValLysPheProGlyLysGlyGlnIleValGlyGlyValTyrLeuLeuThrArgArg  40
QY      120 GGGCCCCAGTTGGGTGTGCGTCAGTCAGTCCGCAAGACTTCGACGCGGTCCCAACTGCAGT  179
Db      41 GlyProIArgLeuGlyValAlaArgAlaThrArgLysThrSerGluArgSerGlnProArgGly  60
QY      180 AAGCGCCACCAACCCATCCCGCGCGCGCCGACCAACCGAGGCGCAGTCTCGGCTCAGCCGCGG  239
Db      61 ArgArgGlnProIleProLysAlaAlaArgArgProGlnGlyAlaArgAlaIleAlaGlnProIle  80
QY      240 TACCTTGGCCCTATATGGGAATGAGGCGTGCGGGTGGCGAGGAGGTGCTCTGTCGCCCG  299
Db      81 TyrProIleProLeuTyrGlyLysnGlnGlyLeuGlyTyrIleAlaGlyTyrPheLeuLeuSerPro  100
QY      300 CGCGGCTCCGCGCGTGTGGTGGCGCCCAATATACCCCGCGCCAGG  344
Db      101 ArgGlySerAlaArgProSerIleProIleProThrAspProAlaArgArg  115

RESULT 8
POLG_HCVHK      STANDARD;      PRT;      520 AA.
AC      001403;
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      01-JUL-1993 (Rel. 40, Last annotation update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Genome polypeptide [Contents: Capsid protein C (Core protein) (p22);
DE      Envelope glycoprotein EI (gp32) (gp35); Envelope glycoprotein E2
DE      (gp68) (GP70) (NS1) (Fragment)."]
OS      Hepatitis C virus (isolate HCV-FF) (HCV).
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
RX      NCBI_TaxID=31644;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=93010930; PubMed=1383400;
RA      Abe K., Inchauspe G., Fujisawa K.;
RT      "Genomic characterization and mutation rate of hepatitis C virus
RT      isolated from a patient who contracted hepatitis during an epidemic
RT      of non-A, non-B hepatitis in Japan."
RL      J. Gen. Virol. 73:2725-2729(1992).
CC      "- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC      LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC      PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC      PROTEIN C AND MRNA.
CC
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC
CC      -----
DR      EMBL, D10687; BA001529.1; -.
DR      PIR, J01925; J01925.
DR      InterPro: IPR002531; HCV_NS1.
DR      InterPro: IPR002522; HCV_capsid.
DR      InterPro: IPR002521; HCV_core.
DR      InterPro: IPR002519; HCV_env.
DR      Pfam: PF01519; HCV_env.1.
DR      Pfam: PF01542; HCV_core.1.
DR      Pfam: PF01543; HCV_capsid.1.
DR      Pfam: PF01560; HCV_NS1.1.
DR      ProDom: PD186062; HCV_NS1.1.

```


OY 60 GAGCTTAAGTTCAGGCGGCGTGCAGATGTTGAGTTCAGTTCAGGCGAGG 119
 DB 21 AspVallyshpethroglyglyglylnilevalglyglylvaltyrleuenduprotrarg 40
 OY 120 GGGCCCCAGTGGGTGGTGGTGCAGTGCAGACCTCCGAGCGGTGCACCTGCAGT 179
 DB 41 GlyProtrgluenglyvalargalathrarglysthrsergluargserglproarggly 60
 OY 180 AGCGCCCAACCCATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 239
 DB 61 ArgArgGlnProleprolyrsparargsertrnglylsysetrpglylsprogly 80
 OY 240 TACCTTGGCCCTATATGGAATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 299
 DB 81 TyrProtrProleuetyrtylglmsnlyglcysglytyrpalaglytyrpleuSerPro 100
 OY 300 CCGCGCTCG 344
 DB 101 ArgGlySerArgProtrProtrProtrProtrProtrProtrProtrProtr 115
 RESULT 10
 POLG_HCVJ8
 ID POLG_HCVJ8 STANDARD: PRT: 3033 AA.
 AC P26661:
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 41, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP58) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitis C)
 DE (EC 3.4.21.98); Nonstructural protein NS4 (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate HC-J8) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11115;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kunita K., Okada S.-I., Yamamoto K., Tanaka T.,
 RA Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RT Virology 188:331-341(1992).
 RL -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA] (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D10988; BAA01761.1; -
 CC DR PIR: A40250; GNMVJ8.
 CC HSSP: P27958; 1HEI.

DR MEROPS: S29.001; -
 DR MEROPS: U39.001; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_NS5b.
 DR InterPro: IPR002522; HCV_NS5c.
 DR InterPro: IPR002521; HCV_NS5d.
 DR InterPro: IPR002519; HCV_NS5e.
 DR Pfam: PF00998; HCV_RdRp; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01006; HCV_NS4d; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF01538; HCV_NS5b; 1.
 DR Pfam: PF01539; HCV_NS5c; 1.
 DR Pfam: PF01542; HCV_NS5d; 1.
 DR Pfam: PF01543; HCV_NS5e; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR Polypol: Glycoprotein; Transferrin; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural
 FT INT_MET 1
 FT CHAIN 1
 FT CHAIN 115
 FT CHAIN 116
 FT CHAIN 191
 FT CHAIN 192
 FT CHAIN 384
 FT CHAIN 733
 FT CHAIN 1011
 FT CHAIN 1010
 FT CHAIN 1620
 FT CHAIN 1866
 FT CHAIN 2017
 FT CHAIN 2018
 FT CHAIN 3033
 FT TRANSMEM 347
 FT ACT_SITE 1087
 FT ACT_SITE 1087
 FT ACT_SITE 1111
 FT ACT_SITE 1111
 FT ACT_SITE 1169
 FT ACT_SITE 1169
 FT N_BIND 1234
 FT N_BIND 1241
 FT SITE 1320
 FT SITE 1323
 FT CARBOHYD 196
 FT CARBOHYD 209
 FT CARBOHYD 233
 FT CARBOHYD 233
 FT CARBOHYD 299
 FT CARBOHYD 299
 FT CARBOHYD 305
 FT CARBOHYD 305
 FT CARBOHYD 417
 FT CARBOHYD 417
 FT CARBOHYD 423
 FT CARBOHYD 423
 FT CARBOHYD 430
 FT CARBOHYD 430
 FT CARBOHYD 448
 FT CARBOHYD 448
 FT CARBOHYD 477
 FT CARBOHYD 477
 FT CARBOHYD 534
 FT CARBOHYD 534
 FT CARBOHYD 542
 FT CARBOHYD 542
 FT CARBOHYD 558
 FT CARBOHYD 558
 FT CARBOHYD 578
 FT CARBOHYD 578
 FT CARBOHYD 627
 FT CARBOHYD 627
 FT CARBOHYD 649
 FT CARBOHYD 649
 FT CARBOHYD 1091
 FT CARBOHYD 1091
 FT CARBOHYD 2038
 FT CARBOHYD 2038
 FT CARBOHYD 2359
 FT CARBOHYD 2359
 FT CARBOHYD 2811
 FT CARBOHYD 2811
 SO SEQUENCE 3033 AA; 330177 MW; 1A173E7E381FDIA CRC64;

Alignment Scores:

Pred. No.: 8.16e-37
 Score: 542.00
 Percent Similarity: 92.17%
 Best Local Similarity: 86.09%
 Query Match: 84.82%
 DB: 1
 Length: 3033
 Matches: 99
 Conservative: 7
 Mismatches: 9
 Indels: 1
 Gaps: 0

```

US-09-873-224-147 (1-345) x POLG_HCVJ8 (1-3033)
OY 1 ATGAGCAGACTTCTTAACCAACAAAGAAAACCAACCAACCC-CGGCCACAG 59
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgArGArgProGln 20
OY 60 GACGTTAAGTCCAGGCGCGGTGAGATCGTGTGGAGCTTACCTGCTACCAACGAG 119
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 21 AspValLysPheProGlyGlyGlnIleValGlyValTyrLeuLeuProArGArg 40
OY 120 GGGCCCCAGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 179
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 41 GlyProArgLeuGlyValAlaGlnArgLysThrSerGlnArgSerGlnProArGArg 60
OY 180 AGGCGCCCAACCCATCCCGAGGCGCGCGCAACCGAGGAGCTCTGGGCTACGCCGGG 239
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 61 ArgArgGlnProIleProLysAspArgArgSerThrGlyLysSerTrpGlyLysProGly 80
OY 240 TACCTTGGCCCTATATGGAGATGAGGCTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 299
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 81 TyrProTrpProLeuTyrGlyAsnGlnGlyGlyTyrPalaglyTrpLeuLeuSerPro 100
OY 300 CGGCGCTCTCGCCGTCGTCGCGCCCAATGACCCCGCGCGCAGG 344
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 101 ArgGlySerArgProThrTrpGlyLysProThrAspProArGHisArg 115

RESULT 11
POLG_HCVJ5 STANDARD: PRT; 737 AA.
ID POLG_HCVJ5
AC p27960;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-J5) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11112;
RN [1]
RP MEDLINE=92230232; Pubmed=1314459;
RA Okamoto H., Kuzai K., Okada S.I., Yamamoto K., Iizuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.,
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: D10075; BAA0096.1; -
DB InterPro: IPR002531; HCV_NSI.
DB InterPro: IPR002522; HCV_capsid.
DB InterPro: IPR002521; HCV_core.
DB InterPro: IPR002519; HCV_env.
DB Pfam: PF01539; HCV_env; 1.
DB Pfam: PF01542; HCV_core; 1.

```

```

DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01560; HCV_NSI; 1.
DR Prodom: PD186062; HCV_NSI; 1.
KW Polypeptide; Glycoprotein; Coat protein; Envelope protein;
KW Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1
FT CHAIN 115
FT CHAIN 191
FT CHAIN 192
FT CHAIN 383
FT CHAIN 384
FT CHAIN 733
FT CHAIN >737
FT TRANSMEM 347
FT CARBOHYD 334
FT CARBOHYD 196
FT CARBOHYD 209
FT CARBOHYD 234
FT CARBOHYD 305
FT CARBOHYD 305
FT CARBOHYD 417
FT CARBOHYD 423
FT CARBOHYD 430
FT CARBOHYD 430
FT CARBOHYD 448
FT CARBOHYD 477
FT CARBOHYD 534
FT CARBOHYD 542
FT CARBOHYD 558
FT CARBOHYD 578
FT CARBOHYD 627
FT CARBOHYD 649
FT CARBOHYD 737
FT NON_TER 737
SQ SEQUENCE 737 AA; 81207 MW; 3AF699D82AD501B1 CMC64;

```

```

Alignment Scores:
Pred. No.: 1.05e-36
Score: 541.00
Percent Similarity: 92.17%
Best Local Similarity: 86.96%
Query Match: 84.66%
DB: 1
Length: 737
Matches: 100
Conservative: 6
Mismatch: 9
Indels: 1
Gaps: 0

```

```

US-09-873-224-147 (1-345) x POLG_HCVJ5 (1-737)
OY 1 ATGAGCAGACTTCTTAACCAACAAAGAAAACCAACCAACCC-CGGCCACAG 59
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgArGArgProGln 20
OY 60 GACGTTAAGTCCAGGCGCGGTGAGATCGTGTGGAGCTTACCTGCTACCAACGAG 119
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 21 AspValLysPheProGlyGlyGlnIleValGlyValTyrLeuLeuProArGArg 40
OY 120 GGGCCCCAGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 179
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 41 GlyProArgLeuGlyValAlaGlnArgLysThrSerGlnArgSerGlnProArGArg 60
OY 180 AGGCGCCCAACCCATCCCGAGGCGCGCGCAACCGAGGAGCTCTGGGCTACGCCGGG 239
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 61 ArgArgGlnProIleProLysAspArgArgSerThrGlyLysSerTrpGlyLysProGly 80
OY 240 TACCTTGGCCCTATATGGAGATGAGGCTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 299
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 81 TyrProTrpProLeuTyrGlyAsnGlnGlyGlyTyrPalaglyTrpLeuLeuSerPro 100
OY 300 CGGCGCTCTCGCCGTCGTCGCGCCCAATGACCCCGCGCGCAGG 344
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 101 ArgGlySerArgProThrTrpGlyLysProThrAspProArGHisArg 115

RESULT 12
POLG_HCVJ6 STANDARD: PRT; 3033 AA.
ID POLG_HCVJ6
AC p26660;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);

```


ID	STANDARD	PRI	3010 AA
AC	POLG_HCV1W		
AC	P29846		
DT	01-APR-1993 (Rel. 25, Created)		
DT	01-APR-1993 (Rel. 25, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Genome polypeptide [contains: Capsid protein C (core protein) (P22);		
DE	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2		
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)		
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)		
DE	(EC 3.4.21.99); Nonstructural protein NS4A (P4); Nonstructural protein		
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein		
DE	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].		
OS	Hepatitis C virus (isolate Taiwan) (HCV).		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepacivirus.		
RN	NCBI_TaxID=31645;		
RN	[1]		
RE	SEQUENCE FROM N.A.		
RA	MEDLINE=92230206; PubMed=1314449;		
RA	Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;		
RT	"The Taiwanese hepatitis C virus genome: sequence determination and		
RT	mapping the 5' termini of viral genomic and antigenomic RNA.";		
RL	Virology 188:102-113(1992).		
CC	-I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE		
CC	HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.		
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.		
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral		
CC	precursor polypeptide, commonly with Asp or Glu in the P6		
CC	position. Cys or Thr in P1 and Ser or Ala in P1'.		
CC	-I- CATALYTIC ACTIVITY: N nucleoside triphosphate +		
CC	(RNA)(N).		
CC	-I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A		
CC	LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:		
CC	PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF		
CC	PROTEIN C AND RNA.		
CC	-I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL: M84754; -I- NOT_ANNOTATED_CDS.		
CC	-----		
DR	PIR: A40244; GMYWTM.		
DR	HSSP: P26663; IXP.		
DR	MEROPS: S29.001; -I-		
DR	InterPro: IPR001410; DEAD.		
DR	InterPro: IPR002531; HCV_NS1.		
DR	InterPro: IPR002518; HCV_NS2.		
DR	InterPro: IPR004109; HCV_NS3.		
DR	InterPro: IPR000745; HCV_NS4A.		
DR	InterPro: IPR001490; HCV_NS4B.		
DR	InterPro: IPR002868; HCV_NS5A.		
DR	InterPro: IPR002166; HCV_RdRp.		
DR	InterPro: IPR002522; HCV_Capsid.		
DR	InterPro: IPR002521; HCV_env.		
DR	InterPro: IPR002519; HCV_core.		
DR	Pfam: PF00998; HCV_RdRp; 1.		
DR	Pfam: PF01001; HCV_NS4B; 1.		
DR	Pfam: PF01006; HCV_NS4A; 1.		
DR	Pfam: PF01506; HCV_NS5A; 1.		
DR	Pfam: PF01538; HCV_NS2; 1.		
DR	Pfam: PF01539; HCV_env; 1.		
DR	Pfam: PF01542; HCV_core; 1.		
DR	Pfam: PF01543; HCV_Capsid; 1.		
DR	Pfam: PF01560; HCV_NS1; 1.		
DR	Pfam: PF02907; HCV_NS3; 1.		
DR	ProDom: PD186062; HCV_NS1; 1.		
FW	Polypeptide: Glycoprotein; Transferase; RNA-directed RNA polymerase;		

KM	Core protein; Coat protein; Envelope protein; Helicase; ATP-binding:
KW	Transmembrane; Nonstructural
FT	INT_MET 1 1
FT	CHAIN 1 115
FT	CHAIN 116 191
FT	CHAIN 192 383
FT	CHAIN 384 729
FT	CHAIN 730 1006
FT	CHAIN 1007 1615
FT	CHAIN 1616 1862
FT	CHAIN 1863 2013
FT	CHAIN 2014 3010
FT	TRANSMEM 347 369
FT	ACT_SITE 1083 1083
FT	ACT_SITE 1107 1107
FT	ACT_SITE 1165 1165
FT	NP_BIND 1230 1237
FT	SITE 1316 1319
FT	CARBOHYD 196 196
FT	CARBOHYD 209 209
FT	CARBOHYD 233 233
FT	CARBOHYD 234 224
FT	CARBOHYD 250 250
FT	CARBOHYD 305 305
FT	CARBOHYD 417 417
FT	CARBOHYD 423 423
FT	CARBOHYD 430 430
FT	CARBOHYD 448 448
FT	CARBOHYD 532 532
FT	CARBOHYD 540 540
FT	CARBOHYD 556 556
FT	CARBOHYD 576 576
FT	CARBOHYD 623 623
FT	CARBOHYD 645 645
FT	CARBOHYD 2041 2041
FT	CARBOHYD 2077 2077
FT	CARBOHYD 2240 2240
FT	CARBOHYD 2529 2529
FT	CARBOHYD 2788 2788
SO	SEQUENCE 3010 AA; 327047 MW; AAD267D5CDE215 CRC64;
 Alignment Scores:	
Pred. No.:	5-276-36
Score:	532.00
Percent Similarity:	92.17%
Best Local Similarity:	86.09%
Query Match:	83.26%
DH:	1 Gaps: 0
 US-09-873-224-147 (1-345) x POLG_HCVTM (1-3010)	
OY	1 ATGAGCACACTTCTCTAAACCAACAAGAAAAACAAAGAACAACGCC-CGGCCACAG 59 Db 1 MetSerThrsmSglyysProglInArglyshThrylsArgsnthrsmAtrgProgn 20
OY	60 GAGCTTAAGTCGCCAGCGCGCGGTCAAGTCTTGTCGGAGATTACGTGTACCAGGCAG 11 Db 21 AspPallyshheProsiyelylsglnlleValisylvalTyrlleuennProArqly 40
OY	120 GGCCCCAGTTGGGTGGTGCGTCGATGCGCAAGACTTCCAGCGGGTGCACACTCGCAGT 17 ::: Db 41 GlyProArqlyenUlyalAlrqlahlrArqlyshTrlpelunrgerelnProArqly 60
OY	180 AGGCGCCAACCCATCCGCCCAAGSGCGCGCGCAACCGAGSGCGAGTCTGGGCTCAGCCC 23 ::: ::: Db 61 ArgArqlynlrrolleProlysAlarqlynProglnglyArqlatrrplnaglnProgly 80
OY	240 TACCCTTGGCCCTTAATAGCGGAATGAGGGCTGGCGGGTGGACAGGTGCTCTGCCCG 29 Db 81 TyrProrlrProleutyrelYasneUlyenUlytrPladlyTrpleuValserPro 100
OY	300 CGGCGCTTCGCCCGTCGCGGGAGGCGCAAATGACCCCGCGCGCAGG 344


```

QY 246 TGGC-----CCCTATATGGGAATGAGGCGCTCGGGTGGCA 281
    |||
    ||| :|||::
Db 266 yglYleproslYalglYalValProglYalValglYalProglYalValglYalVal 286
QY 282 GGGTGGCTCCGTCGCCGGCGGCTCGCGCTCGT-----GG 320
    |||||
    ||||| :|||::
Db 286 lglYlYlProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 306
QY 321 GAGCCAAATGACCCCGCGCAGCA 345
    |||
    ||| :|||::
Db 306 yAlaGlYalValleuProglYalValglYal 314

RESULT 15
SM50_STRPU STANDARD; PRT; 445 AA.
ID SM50_STRPU
AC P11994;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50 kDa spicule matrix protein precursor.
GN SM50.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinozoa; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87162883; PubMed=3030858;
RA Suvov H.M., Benson S.C., Robinson J.J., Britten R.J., Wilt F.H.,
RA Davidson E.H.;
RT "A lineage-specific gene encoding a major matrix protein of the sea
RT urchin embryo spicule. II. Structure of the gene and derived sequence
RT of the protein.";
RL Dev. Biol. 120:507-519(1987).
RN [2]
RP REVISIONS.
RX MEDLINE=91209669; PubMed=2019323;
RA Katoh-Fukui Y., Noce T., Ueda T., Fujiwara Y., Hashimoto N.,
RA Hisashinakagawa T., Killian C.E., Livingston B.T., Wilt F.H.,
RA Benson S.C., Suvov H.M., Davidson R.H.;
RT "The corrected structure of the SM50 spicule matrix protein of
RT Strongylocentrotus purpuratus.";
RL Dev. Biol. 145:201-202(1991).
CC -1- FUNCTION: MAJOR MATRIX PROTEIN OF THE SEA URCHIN EMBRYO SPICULE.
CC THE FUNCTION OF THE MATRIX PROTEINS IS TO DIRECT CRYSTAL GROWTH IN
CC CERTAIN ORIENTATIONS AND INHIBIT GROWTH IN OTHERS.
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE
CC MICROMERE/PRIMARY MESENCHYME CELLS (PMC) LINEAGE.
CC -1- DEVELOPMENTAL STAGE: DETECTED AT HATCHING BLASTULA, IT REMAINS AT
CC A CERTAIN LEVEL FROM THE MESENCHYME BLASTULA TO THE PLUTEUS
CC STAGES.
CC -1- DOMAIN: THE REPETITIVE DOMAIN MAY PROVIDE A CALCITE BINDING
CC MATRIX.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE SM50 FAMILY.
CC
CC -----
CC This SM50-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: M16231; AAA30071.1;
CC EMBL: M16230; AAA30071.1; JOINED.
CC PIR: A27263; A27263.
CC InterPro: IPR001304; Lectin_C.
CC SMART: SM00034; CLECT; 1.
CC PROSITE: PS00615; C_TYPE_LLECTIN_1; FALSE_NEG.
CC PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.

```

```

KW Matrix protein; Repeat; Signal.
FT SIGNAL 1 15
FT CHAIN 16 445 50 KDA SPICULE MATRIX PROTEIN.
FT DOMAIN 29 159 C-TYPE LECTIN.
FT DOMAIN 123 180 PRO-RICH.
SQ SEQUENCE 445 AA; 46262 MW; 10D08FDE088E639A CRC64;

Alignment Scores:
Pred. No.: 1.78 Length: 445
Score: 94.00 Matches: 43
Percent Similarity: 38.05% Conservative: 8
Best Local Similarity: 32.09% Mismatches: 38
Query Match: 14.71% Indels: 46
DB: 1 Gaps: 9

US-09-873-224-147 (1-345) x SM50_STRPU (1-445)
QY 36 AAGAAACACCAACCC-----CGGCCACAGAGCTTAAGTCCAGGCGCGGTGATC 89
    |||
    ||| :|||::
Db 158 LysArgGlnAsnProProValArgProGlyGln-----GlyGlyArgGlnIle 173
QY 90 GTTGTGAGATTACGTGCTACCAACGAGGCGCCCACTTGCGTGTGCTGACG----- 144
    |||||
    ||||| :|||::
Db 174 ProGlnGlyVal-----GlyProGln-TripGlnAlaValGlnValTh 187
QY 145 -----TCCGCAAGACTTCGAGAGCGGTGCAACTGCGAGTAGCGCCAA 188
    |||
    ||| :|||::
Db 187 rAlaMetArgAlaPheValCysGlnValProAlaGlyArgAsnIleProIleGlyGlnI 207
QY 189 CCCA-----TCCGAGGCGCGGCCAAGCAGAGGCG 218
    |||
    ||| :|||::
Db 207 nProGlyMetGlyGlnGlyPheGlyAsnGlnGlnProGlyMetGlyGlyArgGlnPr 227
QY 219 AGGTCTGTGGGCGCAGCGCGGTACCTTGGC-----CCCTA 254
    |||
    ||| :|||::
Db 227 oglyPheGlyAsnGlnProGlyMetGlyGlyArgGlnProGlyPheGlyAsnGlnProG 247
QY 255 TATGGGAATGAGGCGTGCAGGT-----GGCAGGCGTGCCTGTCGCCG 299
    |||||
    ||||| :|||::
Db 247 yMetGlyGlyArgGlnProGlyTTPeGlyAsnGlnProGlyValGlyGlyArgGlnProG 267
QY 300 -----CGCGCTCGCGCGCGTGTGCGGCGCCCAATGACCC 335
    |||
    ||| :|||::
Db 267 yMetGlyGlyGlnGlnProGlyTTPeGly---AsnGlnPro 279

```

Search completed: February 19, 2003, 02:12:06
 Job time : 29 secs



GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: February 19, 2003, 02:07:50 ; Search time 43.5 Seconds

(without alignments)
3268.334 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 639

Sequence: 1 ATGAGCAGACACTTCCTAAGC.....AAATGACCCCGCGCAGAGA 345

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DBV=xlp
-Q/cgn2.1/USPTO.spool/US09873224/rumat.14022003.092705.27450/app.query.fasta.1.519
-DB=SPTRMBL.21 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLION=300 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USFR=US09873224.GCEN.1.1.86.0.rumat.14022003.092705.27450 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NBS_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARM_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SPTRMBL.21:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organella:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
1	604	94.5	415 12	Q68802 hepatitis c

2	604	94.5	415 12	Q68808 hepatitis c
3	598	93.6	415 12	Q81294 hepatitis c
4	596	93.3	415 12	Q68794 hepatitis c
5	591	92.5	415 12	Q68806 hepatitis c
6	590	92.3	415 12	Q68799 hepatitis c
7	590	92.3	3019 12	Q68801 hepatitis c
8	588	92.0	191 12	Q68108 hepatitis c
9	588	92.0	191 12	Q68114 hepatitis c
10	588	92.0	191 12	Q68127 hepatitis c
11	588	92.0	503 12	Q81752 hepatitis c
12	588	92.0	503 12	Q81822 hepatitis c
13	588	92.0	503 12	Q81827 hepatitis c
14	588	92.0	3021 12	Q81258 hepatitis c
15	587	91.9	191 12	Q68125 hepatitis c
16	587	91.9	415 12	Q81554 hepatitis c
17	585	91.5	124 12	Q68894 hepatitis c
18	584	91.4	503 12	Q81489 hepatitis c
19	583	91.2	114 12	Q68893 hepatitis c
20	580	90.8	415 12	Q81546 hepatitis c
21	580	90.8	3021 12	Q68870 hepatitis c
22	579	90.6	114 12	Q68892 hepatitis c
23	578	90.5	123 12	Q68895 hepatitis c
24	577	90.3	411 12	Q81813 hepatitis c
25	576	90.1	191 12	Q91K26 hepatitis c
26	575.5	90.1	415 12	Q81548 hepatitis c
27	575	90.0	415 12	Q81550 hepatitis c
28	574	89.8	115 12	Q68898 hepatitis c
29	574	89.8	191 12	Q68152 hepatitis c
30	574	89.8	191 12	Q68155 hepatitis c
31	574	89.8	191 12	Q68157 hepatitis c
32	574	89.8	191 12	Q91TF7 hepatitis c
33	574	89.8	319 12	Q81263 hepatitis c
34	574	89.8	416 12	Q39647 hepatitis c
35	573	89.7	191 12	Q68576 hepatitis c
36	573	89.7	326 12	Q81257 hepatitis c
37	573	89.7	415 12	Q81272 hepatitis c
38	572	89.5	191 12	Q91K27 hepatitis c
39	572	89.5	191 12	Q91K25 hepatitis c
40	572	89.5	193 12	Q91TF6 hepatitis c
41	571.5	89.4	416 12	Q81265 hepatitis c
42	571	89.4	317 12	Q81260 hepatitis c
43	571	89.4	415 12	Q81552 hepatitis c
44	570.5	89.3	415 12	Q81266 hepatitis c
45	570	89.2	122 12	Q80RJ5 hepatitis c

ALIGNMENTS

RESULT 1
Q68802 PRELIMINARY: PRT: 415 AA.

AC 068802;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70)
DE (NS1)] (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepatitis C virus.
OC Hepatitis C virus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JK055;
RC MEDLINE=96226020; PubMed=8627233;
RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F.,
RA Lesmana L.A., Miyakawa Y., Mayumi M.;
RT *Hepatitis C virus variants from Jakarta, Indonesia classifiable into
RT novel genotypes in the second (2e and 2f), tenth (10a) and eleventh
RT (11a) genetic groups.";
RT J. Gen. Virol. 77:293-301(1996).
CC -!- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL; D49750; BAA08584.1; -.

DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Coat protein: Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 415 415
 SQ SEQUENCE 415 AA; 44937 MW; D90085CB58FF6B CRC64;

Alignment Scores:

Pred. No.:	1,08e-48	Length:	415
Score:	604.00	Matches:	113
Percent Similarity:	99.13%	Conservative:	1
Best Local Similarity:	98.26%	Mismatches:	1
Query Match:	94.52%	Indels:	1
DB:	12	Gaps:	0

US-09-873-224-147 (1-345) x Q68802 (1-415)

QY 1 ATGAGCACACTTCTTAACCAAGAAAACCAACCAACCC-CGGCCACAG 59
 |||||||
 DB 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
 QY 60 GACGTTAAGTCCAGCGCGGCGTCAATCGTTGGAGTTTACGCTCAACAGCAG 119
 |||||||
 DB 21 AspValLysPheProGlyGlyGlyGlnIleValGlyValValLeuProArgArg 40
 QY 120 GGGCCCCAGTTGGGTGGTGGTCAAGTGGCAAGCTTCCAGCGGTGCACTGCAGT 179
 |||||||
 DB 41 GlyProLysLeuGlyValAlaValAlaValArgLysThrSerGlnArgSerGlnProArgSer 60
 QY 180 AGGCGCAACCCATCCCGAGGCGCGGCGCAACCGGAGGCGAGGCTGCTGCGTCCAGCGGG 239
 |||||||
 DB 61 ArgArgGlnProIleProArgAlaArgArgThrGlnGlyArgSerThrPalaglnProGly 80
 QY 240 TACCTTGGCCCTATATGGAATGAGGCGTGGGCGGAGGAGGCTGCTGCTGCTCCCG 299
 |||||||
 DB 81 TyrProTrpProLeuTyrGlyAsnGlnGlyGlyGlyTrpAlaglnTrpLeuLeuSerPro 100
 QY 300 CGCGGCTCTCGCCGCTGCGTGGGCGCAATGACCCCGCGGCGAG 344
 |||||||
 DB 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 2

ID 068808 PRELIMINARY: PRT: 415 AA.

AC 068808;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Genome polyprotein [contains: envelope glycoprotein E2 (GP68) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JK072;
 RX MEDLINE=36226020; PubMed=6627233;
 RA Tokita H., Okamoto H., Itzuka H., Kishimoto J., Tsuda F.,
 RA Lesmana L.A., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Jakarta, Indonesia classifiable into
 RT novel genotypes in the second (2e and 2f), tenth (10a) and eleventh
 RT (11a) genetic groups";
 J. Gen. Virol. 77:293-301(1996).
 CC - - SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
 DR EMBL: D49753; BAA06587.1; -;
 DR InterPro: IPR002522; HCV_capsid.

DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Coat protein: Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 415 415
 SQ SEQUENCE 415 AA; 44962 MW; 3CA7D25EA4C1869 CRC64;

Alignment Scores:

Pred. No.:	1,08e-48	Length:	415
Score:	604.00	Matches:	113
Percent Similarity:	99.13%	Conservative:	1
Best Local Similarity:	98.26%	Mismatches:	1
Query Match:	94.52%	Indels:	1
DB:	12	Gaps:	0

US-09-873-224-147 (1-345) x Q68802 (1-415)

QY 1 ATGAGCACACTTCTTAACCAAGAAAACCAACCAACCC-CGGCCACAG 59
 |||||||
 DB 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
 QY 60 GACGTTAAGTCCAGCGCGGCGTCAATCGTTGGAGTTTACGCTCAACAGCAG 119
 |||||||
 DB 21 AspValLysPheProGlyGlyGlyGlnIleValGlyValValLeuProArgArg 40
 QY 120 GGGCCCCAGTTGGGTGGTGGTCAAGTGGCAAGCTTCCAGCGGTGCACTGCAGT 179
 |||||||
 DB 41 GlyProLysLeuGlyValAlaValAlaValArgLysThrSerGlnArgSerGlnProArgSer 60
 QY 180 AGGCGCAACCCATCCCGAGGCGCGGCGCAACCGGAGGCGAGGCTGCTGCGTCCAGCGGG 239
 |||||||
 DB 61 ArgArgGlnProIleProArgAlaArgArgThrGlnGlyArgSerThrPalaglnProGly 80
 QY 240 TACCTTGGCCCTATATGGAATGAGGCGTGGGCGGAGGAGGCTGCTGCTGCTCCCG 299
 |||||||
 DB 81 TyrProTrpProLeuTyrGlyAsnGlnGlyGlyGlyTrpAlaglnTrpLeuLeuSerPro 100
 QY 300 CGCGGCTCTCGCCGCTGCGTGGGCGCAATGACCCCGCGGCGAG 344
 |||||||
 DB 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 3

ID 081294 PRELIMINARY: PRT: 119 AA.

AC 081294;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Polyprotein (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FD-3/93;
 RX MEDLINE=9518942; PubMed=7883898;
 RA Hotta H., Handajani R., Ingelusida M., Soemarto W., Doi H.,
 RA Miyajima H., Homma M.;
 RT "Subtype analysis of hepatitis C virus in Indonesia on the basis of
 RT NS5B region sequences";
 J. Clin. Microbiol. 32:3049-3051(1994).
 DR EMBL: D30047; BAA06283.1; -;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 FT NON_TER 119 119
 SQ SEQUENCE 119 AA; 13620 MW; 471715D6F84E58C8 CRC64;

Alignment Scores:

Pred. No.: 3.85e-48 Length: 119
 Score: 598.00 Matches: 112
 Percent Similarity: 98.26% Conservative: 1
 Best Local Similarity: 97.39% Mismatches: 2
 Query Match: 93.58% Indels: 1
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x 081294 (1-119)

```

QY 1 ATGAGCACACTTCTTAACCAAGAAAGAAAAAGAACACCAACC-CGGCCACAG 59
    |||||||
Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGln 20
QY 60 GACGTTAGTTCCAGGCGCGGTGAGATCGTTGAGATTAGCTGTACGCTACGACGAG 119
    |||||||
Db 21 AspValLysPheProGlyGlyGlnIleValIGlyValValValLeuProArgArg 40
QY 120 GGGCCCGAGTTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 179
    |||||||
Db 41 AlaProArgLeuGlyValAlaGalaValAlaGlyThrSerGluArgSerGlnProArgSer 60
QY 180 AGCGCCCAACCCATCCCAAGGCGCGCCCAAGCCAGAGGAGGAGGAGGAGGAGGAGGAGG 239
    |||||||
Db 61 ArgArgGlnProIleProArgAlaArgArgThrGlnGlyArgSerTrpAlaGlnProGly 80
QY 240 TACCTTGGCCCTATATGGAATGAGGCTCGGGTGGGAGGGTGGCTCTCTCCCG 299
    |||||||
Db 81 TyrProTrpProLeuTrpLysnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
QY 300 CGCGGCTCTCGCCGCTGCGGCGCCCAATGACCCCGCGGAGG 344
    |||||||
Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

```

RESULT 4

```

ID Q68794 PRELIMINARY: PRT: 415 AA.
AC Q68794:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70)
DE (NS1)] (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JK030;
RX MEDLINE=96226020; PubMed=8627233;
RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F.,
RA Lesmana L.A., Miyakawa Y., Mayumi M.,
RT "Hepatitis C virus variants from Jakarta, Indonesia classifiable into
RT novel genotypes in the second (2e and 2f), tenth (10a) and eleventh
RT (11a) genetic groups."
RL J. Gen. Virol. 77:293-301(1996).
CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: D49747; BAA08581.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 415
SQ SEQUENCE 415 AA: 44797 MW: B97478725A2B6238 CRC64;

```

Alignment Scores:

Pred. No.: 6.14e-48 Length: 415
 Score: 596.00 Matches: 112
 Percent Similarity: 98.26% Conservative: 1
 Best Local Similarity: 97.39% Mismatches: 2
 Query Match: 93.27% Indels: 1
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x 068794 (1-415)

```

QY 1 ATGAGCACACTTCTTAACCAAGAAAGAAAAAGAACACCAACC-CGGCCACAG 59
    |||||||
Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGln 20
QY 60 GACGTTAGTTCCAGGCGCGGTGAGATCGTTGAGATTAGCTGTACGCTACGACGAG 119
    |||||||
Db 21 AspValLysPheProGlyGlyGlnIleValIGlyValValValLeuProArgArg 40
QY 120 GGGCCCGAGTTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 179
    |||||||
Db 41 GlyProLysLeuGlyValAlaGalaValAlaGlyThrSerGluArgSerGlnProArgSer 60
QY 180 AGCGCCCAACCCATCCCAAGGCGCGCCCAAGCCAGAGGAGGAGGAGGAGGAGGAGGAGG 239
    |||||||
Db 61 ArgArgGlnProIleProArgAlaArgArgThrGlnGlyArgSerTrpAlaGlnProGly 80
QY 240 TACCTTGGCCCTATATGGAATGAGGCTCGGGTGGGAGGGTGGCTCTCTCCCG 299
    |||||||
Db 81 TyrProTrpProLeuTrpLysnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
QY 300 CGCGGCTCTCGCCGCTGCGGCGCCCAATGACCCCGCGGAGG 344
    |||||||
Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

```

RESULT 5

```

ID Q68806 PRELIMINARY: PRT: 415 AA.
AC Q68806:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70)
DE (NS1)] (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JK070;
RX MEDLINE=96226020; PubMed=8627233;
RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F.,
RA Lesmana L.A., Miyakawa Y., Mayumi M.,
RT "Hepatitis C virus variants from Jakarta, Indonesia classifiable into
RT novel genotypes in the second (2e and 2f), tenth (10a) and eleventh
RT (11a) genetic groups."
RL J. Gen. Virol. 77:293-301(1996).
CC -I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
DR EMBL: D49752; BAA08586.1; -.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NSI; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT NON_TER 415
SQ SEQUENCE 415 AA: 44979 MW: 3368538D667DEA39 CRC64;

```

Alignment Scores:

Pred. No.: 1.82e-47 Length: 415

DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; HCV_RdRP; 1.
 DR PRODOM: PD186062; HCV_NS1; 1.
 DR PROSITE: PS00599; AA_TRANSFERRIN; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Nonstructural protein; Polypeptide; RNA-directed RNA polymerase;
 KW Transmembrane.
 SQ SEQUENCE 3019 AA; 328210 MW; AF7A6774BCD95FA CRC64;

Alignment Scores:
 Pred. No.: 2,38e-47 Length: 3019
 Score: 590.00 Matches: 111
 Percent Similarity: 97.39% Conservative: 1
 Best Local Similarity: 96.52% Mismatches: 3
 Query Match: 92.33% Indels: 1
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x Q68801 (1-3019)

QY 1 ATGAGCACACTTCTTAACCAAGAAAAAGAAACCAACCC-CGGCCACAG 59
 |||||
 Db 1 MetSerThrLeuProLysProGlnArgIleThrLysArgSnlLeaSnArgProGln 20

QY 60 GACGTTAAGTTCACAGCGCGGTCACATGCTGGTACGTTACGTCACACGAG 119
 |||||
 Db 21 AspValLysPheProGlyGlyGlnIleValGlyValIleLeuProArgArg 40

QY 120 GCGCCCGAGTGGGTGTCGTCAGTGCAGAGCTTCGAGCGGTGCAACCTCGCAGT 179
 |||||
 Db 41 GlyProLysLeuGlyValArgAlaValAlaGlyThrSerGluArgSerGlnProArgSer 60

QY 180 AGGCGCCACCCATCCACAGGCGCGCCGGAACCGAGGCGAGTCTCGGCTCAGCCGGG 239
 |||||
 Db 61 ArgArgGlnProIleProIleArgAlaArgThrGlnIleValArgSerTrpAlaGlnProGly 80

QY 240 TACCCCTGGCCCTATATGGAATGAGGAGTGGGCTGGGAGGAGTGGCTCTGTCGCCG 299
 |||||
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100

QY 300 CGCGGCTCTCGCCGTCGTGGGGCCCAATGACCCCGCGCAGG 344
 |||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 8
 Q68108 PRELIMINARY; PRT; 191 AA.
 AC Q68108;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Polypeptide (Fragment).
 GN C.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DK12;
 RX MEDLINE=94336721; PubMed=8058787;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "Sequence analysis of the core gene of 14 hepatitis C virus
 RT genotypes";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243(1994).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=DK12;
 RX MEDLINE=92279243; PubMed=1317578;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "Sequence analysis of the 5' noncoding region of hepatitis C virus";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DK12;
 RX MEDLINE=93376778; PubMed=8396266;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "At least 12 genotypes of hepatitis C virus predicted by sequence
 RT analysis of the putative E1 gene of isolates collected worldwide";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).
 DR EMBL: U10191; AAA21030.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR NON_TER 191 191
 SQ SEQUENCE 191 AA; 20807 MW; A31BD939D5E7B8E2 CRC64;

Alignment Scores:
 Pred. No.: 3.41e-47 Length: 191
 Score: 588.00 Matches: 108
 Percent Similarity: 96.52% Conservative: 3
 Best Local Similarity: 93.91% Mismatches: 4
 Query Match: 92.02% Indels: 1
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x Q68108 (1-191)

QY 1 ATGAGCACACTTCTTAACCAAGAAAAAGAAACCAACCC-CGGCCACAG 59
 |||||
 Db 1 MetSerThrLeuProLysProGlnArgIleThrLysArgSnlLeaSnArgProGln 20

QY 60 GACGTTAAGTTCACAGCGCGGTCACATGCTGGTACGTTACGTCACACGAG 119
 |||||
 Db 21 AspValLysPheProGlyGlyGlnIleValGlyValIleLeuProArgArg 40

QY 120 GCGCCCGAGTGGGTGTCGTCAGTGCAGAGCTTCGAGCGGTGCAACCTCGCAGT 179
 |||||
 Db 41 GlyProLysLeuGlyValArgAlaValAlaGlyThrSerGluArgSerGlnProArgSer 60

QY 180 AGGCGCCACCCATCCACAGGCGCGCCGGAACCGAGGCGAGTCTCGGCTCAGCCGGG 239
 |||||
 Db 61 ArgArgGlnProIleProLysAlaArgArgSerGlnIleValArgSerTrpAlaGlnProGly 80

QY 240 TACCCCTGGCCCTATATGGAATGAGGAGTGGGCTGGGAGGAGTGGCTCTGTCGCCG 299
 |||||
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100

QY 300 CGCGGCTCTCGCCGTCGTGGGGCCCAATGACCCCGCGCAGG 344
 |||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 9
 Q68114 PRELIMINARY; PRT; 191 AA.
 AC Q68114;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Polypeptide (Fragment).
 GN C.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HK10;
 RX MEDLINE=94336721; PubMed=8058787;
 RA Bukh J., Purcell R.H., Miller R.H.;

"Sequence analysis of the core gene of 14 hepatitis C virus genotypes.";
 RT Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243(1994).
 RN [2]
 RC STRAIN=SS2;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92279243; PubMed=1317578;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "Sequence analysis of the 5' noncoding region of hepatitis C virus."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HK10;
 RX MEDLINE=93376778; PubMed=8396266;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).
 DR EMBL: U010197; AAA21036.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 FT NON_TER 191
 SQ SEQUENCE 191 AA; 20807 MW; A31BD939D5E7B8E2 CRC64;

Alignment Scores:
 Pred. No.: 3,41e-47 Length: 191
 Score: 588.00 Matches: 108
 Percent Similarity: 96.52% Conservative: 3
 Best Local Similarity: 93.91% Mismatches: 4
 Query Match: 92.02% Indels: 1
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x Q68114 (1-191)

QY 1 ATGAGCAGCACTTCCCTAAACCAAGAAAAAAGAAACACCA-CCCGGCCACAG 59
 |||||||
 Db 1 MetSerThrLeuProLysProGlnThrLysThrLysArgSerThrLleArgArgProGln 20

QY 60 GACGTTAAGTCCAGCGCGGCGGTGAGATCGTTGAGTTTACGTACACGACAG 119
 |||||||
 Db 21 AspValLysPheProGlyGlyGlnIleValGlyValIleProArgArg 40

QY 120 GGGCCCACTGGGTGGTGGCGGACGACGACGACGCGGTGCAACCTCGCAGT 179
 |||||||
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60

QY 180 AGCGCCCAACCCATCCCGACGCGCGCGAAGCGAGCGGCGTCCGACGCGCGG 239
 |||||||
 Db 61 ArgArgGlnProIleProLysAlaArgArgSerGlnLysArgSerTrpAlaGlnProGly 80

QY 240 TACCTTGCGCCCTATATGGGAATGAGGCGTGGCGGTGGCGAGGCGTCTGTCGCCG 299
 |||||||
 Db 81 TyrProTrpProLeuThrGlyAsnGlnGlyCysGlyTrpIleGlyTrpLeuLeuSerPro 100

QY 300 CGCGGCTCTGCGCGCGTGGGGCCCAATGACCCCGCGCGCAGG 344
 |||||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 10
 Q68127
 ID Q68127 PRELIMINARY: PRT: 191 AA.
 AC Q68127;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DE Polyprotein (Fragment).
 GN
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SS2;
 RX MEDLINE=94336721; PubMed=8058787;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "Sequence analysis of the core gene of 14 hepatitis C virus genotypes."
 RL Proc. Natl. Acad. Sci. U.S.A. 91:8239-8243(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SS2;
 RX MEDLINE=92279243; PubMed=1317578;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "Sequence analysis of the 5' noncoding region of hepatitis C virus."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SS2;
 RX MEDLINE=93376778; PubMed=8396266;
 RA Bukh J., Purcell R.H., Miller R.H.;
 RT "At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238(1993).
 DR EMBL: U010210; AAA21049.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 FT NON_TER 191
 SQ SEQUENCE 191 AA; 20793 MW; A31BD939D5F528E2 CRC64;

Alignment Scores:
 Pred. No.: 3,41e-47 Length: 191
 Score: 588.00 Matches: 108
 Percent Similarity: 96.52% Conservative: 3
 Best Local Similarity: 93.91% Mismatches: 4
 Query Match: 92.02% Indels: 1
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x Q68127 (1-191)

QY 1 ATGAGCAGCACTTCCCTAAACCAAGAAAAAAGAAACACCA-CCCGGCCACAG 59
 |||||||
 Db 1 MetSerThrLeuProLysProGlnThrLysThrLysArgSerThrLleArgArgProGln 20

QY 60 GACGTTAAGTCCAGCGCGGCGGTGAGATCGTTGAGTTTACGTACACGACAG 119
 |||||||
 Db 21 AspValLysPheProGlyGlyGlnIleValGlyValIleProArgArg 40

QY 120 GGGCCCACTGGGTGGTGGCGGACGACGACGACGCGGTGCAACCTCGCAGT 179
 |||||||
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60

QY 180 AGCGCCCAACCCATCCCGACGCGCGCGAAGCGAGCGGCGTCCGACGCGCGG 239
 |||||||
 Db 61 ArgArgGlnProIleProLysAlaArgArgSerGlnLysArgSerTrpAlaGlnProGly 80

QY 240 TACCTTGCGCCCTATATGGGAATGAGGCGTGGCGGTGGCGAGGCGTCTGTCGCCG 299
 |||||||
 Db 81 TyrProTrpProLeuThrGlyAsnGlnGlyCysGlyTrpIleGlyTrpLeuLeuSerPro 100

QY 300 CGCGGCTCTGCGCGCGTGGGGCCCAATGACCCCGCGCGCAGG 344
 |||||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 11
 Q681752
 ID Q681752 PRELIMINARY: PRT: 503 AA.
 AC Q681752;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DE Genome polyprotein (Contains: envelope glycoprotein E2 (GP68) (GP70))

Alignment Scores:	
Pred. No.:	3.5e-47
Score:	568.00
Percent Similarity:	96.528
Best Local Similarity:	93.918
Query Match:	92.026
Ob:	12
	1
	0
Length:	503
Matches:	108
Conservative:	
Mismatches:	4
Indels:	1
Gaps:	0

OY	1	ATGAGACACCTTCCTTAAMCCAAAGAAAACCATAAAGAACAACCCA--CCCCGGCCACG	59
Dd	1	MeSeSerThnLeuProIysPProGIInArLysThrLysArgAsnThrLeaArgRProGIIn	20
OY	60	GACGTAAAGTTCCACAGCGCGGGTCAAGATCGTTGGTGAGATTACGTGTACACGCACAG	119
Dd	21	ASPValLYsPhneProGLyGLyGLInHLevalGLyGLYValTYeValLeuProArgArg	40
OY	120	GGCCCCCAGTTGGGCGGCCGCGACGTCCGAACACTCCGACGCGTCCGAACCTTGCACT	178
Dd	41	GLYPRoArgLeuGLyValAlArgAlaThrAlrLYsThmSerGIInArgSerGIInPrArgGLy	60
OY	180	AAGCCCCAACCCATCCCAAGGCGCGCCGCAACGACGAGGGCAGGCTCGAGCTCACGCCGG	238
Dd	61	ARGAArgGIInProILeProLYsAlARgArgSerGIInGLYArgSerTRPalAGInProGLy	80
OY	240	TACCCCTTAGCCCCTATATGTGGAAATGAGGGGTGGCGGTGGCGAGGTGCTCTGCTCCCG	298
Dd	81	TYrProTrIProLeuTYrGLYAsnGLInGLYcYSGLYTTPALaGLYTTPLeuLeuSerPro	100
OY	300	CGGGGCTCTGCCCCGCTGCTGGGGCCCAAATGACCCCCGGCGCAGG	344
Dd	101	ARGLYSerARgProSerITripGLYProASnASPProARgArg	115
RESULT	12		

Alignment Scores:	
Pred. No.:	3.5e-47
Score:	588.00
Percent Similarity:	96.52%
Best Local Similarity:	93.91%
Query Match:	92.02%
DB:	12
Gaps:	0
Indels:	1
Mismatches:	4
Conservative:	3
Matches:	108
Length:	503

QY	1	ATAGAGCACTTCCTTAACCCAAAGAAAAACCAAGAAACCCAA-CCCGGGCCACAG	59
Db	1	MetSerThrLeuProLysProGlnArgLysThrLysArgGlnThrLysArgProGln	20
QY	60	GAGCTAAATTCCACGGCGGGCTGACATGCTGGTGAGATTACGCTACACGCAGG	119
Db	21	AspValLysPheProGlyGlyGlyGlnIleValGlyValIleValLeuProArgArg	40
QY	120	GGCCCCAGTTGGGTGTCGTCGACGCGCAACACTCCGAGCGGCGAACCTCGCAGT	179
Db	41	GlyProAlaGluGlyValAlaArgAlaThrArgLysThrSerGlnArgSerGlnProArgLys	60
QY	180	AGCGCGCAACCCATCCCGACGGCGCGCGCAACCGAGGGAGGTCCTGGGGTCACCGCGG	239
Db	61	ArgAlaArgGlnProIleProLysAlaIleArgArgSerGlnGlyArgSerThrPalaGlnProGly	80
QY	240	TACCGCTGGCCCAATATGGAATGAGAGCTGGCGGTGGGCAAGGCTGCTCTGCCCG	299

Dd	TyrlptirProLeutylrGlysanGlueGlyCysGlyTTPralagLTyrPleuenseuSerPro	100
Qy	81 TyrlptirProLeutylrGlysanGlueGlyCysGlyTTPralagLTyrPleuenseuSerPro	344
Dd	300 GCGGGCTCTCGCCCGTCGTGGGCGCCCAATGCACCCGGCGCAGG	115
Qy	101 ArgIglySerArgrProserTerTrpgLyProAsnspPrroAArgarG	115
RESULT 13		
ID	Q81827	PRELIMINARY; PRT; 503 AA.
AC	Q81827;	
DT	01-NOV-1996 (TREMBLrel_01, Created)	
DT	01-NOV-1996 (TREMBLrel_01, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel_19, last annotation update)	
DE	Genome polyprotein [contains: envelope glycoprotein E2 (GP68) (GP70) (NS1) (Fragment).]	
OS	Hepatitis C virus.	
OC	Vituses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Hepadnavirus.	
OX	NCBI_TaxID=11103;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-US114;	
FX	MEDLINE=94065664; Pubmed=7504073;	
RA	Okamoto H., Tokita H., Sakamoto M., Horikita M., Kojima M., Iizuka H.,	
RA	Mishiro S.;	
RI	"Characterization of the genomic sequence of type V (or 3a) hepatitis	
RL	C virus isolates and PCR primers for specific detection.";	
RL	J. Gen. Virol. 74:2385-2390(1993).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-US114;	
RA	Okamoto H.;	
RL	Submitted (FEB-1994) to the EMBL/GenBank/DDAJ databases.	
CC	-I- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN EI.	
DR	EMBL: D14309; BAA03255.1; -.	
DR	InterPro: IPR002522; HCV_capsid.	
DR	InterPro: IPR002521; HCV_core.	
DR	InterPro: IPR002519; HCV_env.	
DR	InterPro: IPR002531; HCV_NS1.	
DR	Pfam: PF01543; HCV_capsid; 1.	
DR	Pfam: PF01542; HCV_core; 1.	
DR	Pfam: PF01539; HCV_env; 1.	
DR	Pfam: PF01560; HCV_NS1; 1.	
KW	Prodom, PD186062; HCV_NS1; 1.	
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;	
KW	Polyprotein; Transmembrane.	
FT	CHAIN 1 >191 C.	
FT	CHAIN 192 >383 E.	
FT	CHAIN 384 >503 E2/NS1.	
FT	NON_TER 503	
SQ	SEQUENCE 503 AA; 54833 MW; 31614329C22BDFEE CRC64;	
Alignment Scores:		
Pred. No. :	3.5e-47 Length: 503	
Score:	588.00 Matches: 108	
Percent Similarity:	96.52% Conservative: 3	
Best Local Similarity:	93.91% Mismatches: 4	
Query Match:	92.02% Indels: 1	
DB:	12 Gaps: 0	
US-09-873-224-147 (1-345) x Q81827 (1-503)		
Qy	1 ATGAGCACACTTCCTTAACCAAGAATAAACCAAAAACAACCACCA-CGCCGGCACAG	59
Dd	1 MetSerThrIleuProIyrosproIinArGlyStrHysArGaSnThrIleArgArGProGln	20
Qy	60 GACGTTAAgTTCcCAGGcGcGcgGTcAGATcGTTGtTGAgGTTAcGTCtACACGcAGG	119
Dd	21 AspValIlyshpheProclYelYglYgInIleValIgIyGLyValTYrValIleuProArGarg	40
Qy	120 GGCCCCAGTGGGTGGGTGGGTGCAGGCGCAGACACTTCGAGCGGTGCACACTCGCAGT	179
Dd	41 GLYPtoArGluenGLyAlArGalArIathArGlyStrHiserGIuArGserChnProArGly	60

Oy	180	AGCGCCACCATTCGCCAGGGGCGCGCGAACCGAGGCGAGGCTCGGAGTCAAGCCGG	239
Dd	61	ATGATGlnProIleProLysIlaArgSerSeriUgLYArSerTrpAlaGlnProgly	80
Oy	240	TACCCTTGAGCCCCCTAAATGGAAGAAGAGCGCTGCAGGTGGCAGAGGTGCCTCTCCCG	299
Dd	81	TyrProTrpProLeutrycLYAsnUngLYcysLYTrpAlaGLYTrpLeuSenrpro	100
Oy	300	CGCGGCTCTGCCCCGCTGCGGGGCCCAAATGACCCCCGGCGCAGG	344
Dd	101	ArgLYSerArGProSerTrpLYProAsnAspArGArGArg	115
RESULT 14			
ID	081258	PRELIMINARY:	PRT: 3021 AA.
AC	Q81258;		
DI	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DF	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Genome polypeptide [contains: envelope glycoprotein E2 (GP68) (GP70) (NS1)]		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
CC	Hepadnavirus.		
OX	NCBI_TaxID=11103;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NZLL;		
RA	Sakamoto M.;		
RL	Submitted (SEP-1993) to the EMBL/Genbank/DBJ databases.		
RU	[2]		
RC	STRAIN=NZLL;		
RA	Sakamoto M.;		
RL	Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.		
CU	- I - SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.		
DR	EMBL: D17763; BAA04609.1: "-		
DR	HSSP: P27958; IHET.		
DR	MEROPS: S29.001: -		
DR	MEROPS: U39.001: -		
DR	InterPro: IPR001410; DEAD.		
DR	InterPro: IPR002522; HCV_capsid.		
DR	InterPro: IPR002521; HCV_core.		
DR	InterPro: IPR002519; HCV_env.		
DR	InterPro: IPR002513; HCV_NS1.		
DR	InterPro: IPR002518; HCV_NS2.		
DR	InterPro: IPR004109; HCV_NS3.		
DR	InterPro: IPR000745; HCV_NS4.		
DR	InterPro: IPR001490; HCV_NS4b.		
DR	InterPro: IPR002868; HCV_NS5a.		
DR	InterPro: IPR002166; HCV_RdRp.		
DR	InterPro: IPR002016; peroxidase.		
DR	pfam: PF01543; HCV_capsid_1.		
DR	pfam: PF01542; HCV_core; 1.		
DR	pfam: PF01539; HCV_env; 1.		
DR	pfam: PF01560; HCV_NS1; 1.		
DR	pfam: PF01538; HCV_NS2; 1.		
DR	pfam: PF02907; HCV_NS3; 1.		
DR	pfam: PF01006; HCV_NS4a; 1.		
DR	pfam: PF01001; HCV_NS4b; 1.		
DR	pfam: PF00956; HCV_NS4b; 1.		
DR	pfam: PF00998; HCV_RdRp; 1.		
DR	ProDom: PD186062; HCV_NS1; 1.		
DR	PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.		
KW	ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;		
KM	Nonstructural protein; Polypeptide; RNA-directed RNA polymerase;		
KW	Transmembrane.		
FT	CHAIN	1	C.
FT	CHAIN	192	E1.
FT	CHAIN	384	E2/NS1.
FT	CHAIN	735	NS2.
FT	CHAIN	736	NS3.
FT	CHAIN	1013	NS3.

FT	CHAIN	1664	1717	NS4A.
FT	CHAIN	1718	1978	NS4B.
FT	CHAIN	1979	2430	NS5A.
FT	CHAIN	2431	3021	NS5B.
SQ	SEQUENCE	3021 AA:	329574 MM:	38712CCBC0C19562 CRC64

Alignment Scores:

Fried. NO.:	3.67e-47	Length:	3021
Score:	568.00	Matches:	108
Percent Similarity:	96.52%	Conservative:	3
Best Local Similarity:	93.91%	Mismatches:	4
Query Match:	92.02%	Indels:	1
DB:	12	Gaps:	0

US-09-873-224-147 (1-345) x Q81258 (1-3021)

OY	I	ATGAGCACACTTCCCTAAACCAACAAGAAAAAACCAAAAGAACCAACA - CCGCGGCACAG	59
Dd	1	MetserThirneuproIsProGlnArgIstHrIysArGsnHrILearArgProGln	20
OY	60	GACGTACGTTCCAGCGGGCGGTACAGTCGTTGTCAGATTACGTCGTACACGACAG	118
Dd	21	AspAllylspherosiylElylGlnIleValGlylylValtyrValLeuProArgArg	40
OY	120	GGCCCCAGTTGGGTGGTGTCACATGGCGCAAGACTCCGAGGGTCGACCTGCAGT	179
Dd	41	GlyProArgLeuGlyValAlaGlnAhThrArgIstHrSerClnArgSerClnProArgIgly	60
OY	180	AGCGGCCAACCATCCCAAGGGCGCGCCGAACGACGAGGGCAGCTCCTGGGCTACGCCGGG	239
Dd	61	ArgArgGlnProIleProIlyAlaAlaArgArgSerClnGlyVArgSerTrpAlaGlnProGly	80
OY	240	TACCCTTGCCCCCTATATGGGAATGAGGGCTGGGGGTGGGCAAGGCTGCTCTGTCCCG	299
Dd	81	TyrProTrpProLeuTfryelYasnGlnGlyCySgLYTPRPaLgLYTPRLeuLeuSerPro	100
OY	300	CGGGGCTCTGCCGCTCGTGGGGCCCAATGACGCCCGCGGACAG	344
Dd	101	ArgGlySerArgProSerTerPrGlyProAsnAspProArgArgArg	115

ID	Q68125	PRELIMINARY;	PRT;	191	AA.
AC	Q68125;				

D1	01-NOV-1996	(TIREMBLrel. 01, Created)
DT	01-NOV-1996	(TIREMBLrel. 01, last sequence update)
DT	01-DEC-2001	(TIREMBLrel. 19, last annotation update)
DE	Polypeptide (Fragment).	
NN		

Hepatitis C virus.

CC Hepacivirus, no DNA stage; Flaviviridae,
 NCBI TaxID-11103.
 OX

RN	[1]
RD	SEQUENCE FROM N

RC STRAIN=S2;
BY MEDTNE-04336704

Bukh J., Purcell R.H., Miller R.H.
#Covered

genotypes." ;
hepatitis C virus

[2] 201: 0:0:0: 21:0239-8243 (1994).

```
STRAIN=S2;
```

Bukh J., Purcell R.H., Miller R.H.

Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946(1992).

SEQUENCE FROM N.A.

MEDLINE=93376778; PubMed=8396266;

WILLER R.H., MILLER R.H.,

RT "At least 12 genotypes of hepatitis C virus predicted by sequence
RI analysis of the putative E1 gene of isolates collected worldwide."
RL Proc. Natl. Acad. Sci. U.S.A. 90:8234-8238 (1993).
DR EMBL: U01028; AAA21047.1; -.

	ncv_capsid
DR InterPro; IPR002521;	ncv_core.

Pfam; PF01542; HCV_core; 1.

NON_LEN	SEQUENCE	191 AA:	20821 MW:	98785B3FCFC952AFB CBQ6A:
191	191			

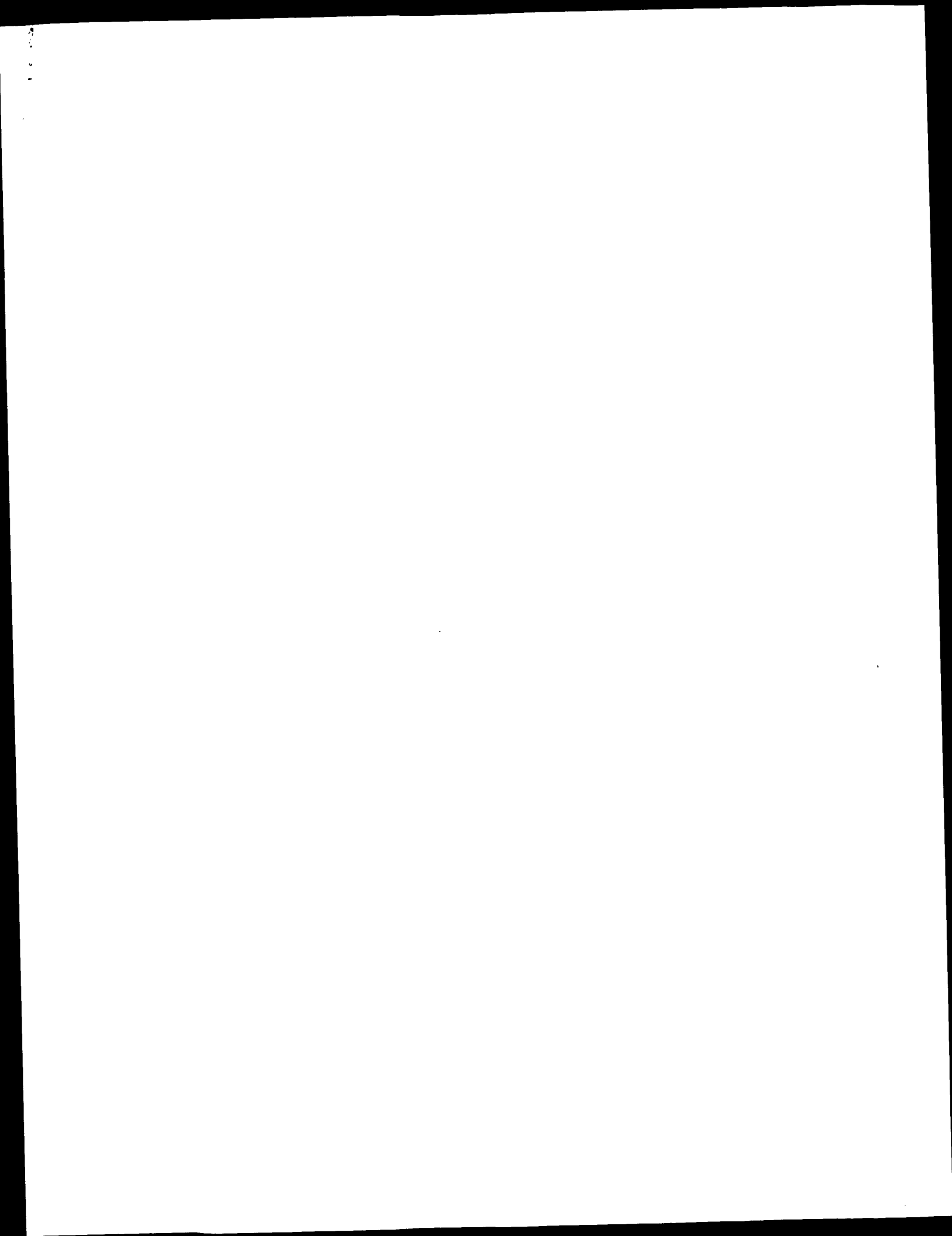
Alignment Scores

File: NO.:	4.33e-47	Length:	19
Score:	587.00	Matches:	107
Percent Similarity:	96.528	Conservative:	1
Best Local Similarity:	93.048	Mismatches:	4
Query Match:	91.868	Indels:	1
DB:	12	Gaps:	0

US-09-873-224-147 (1-345) x Q68125 (1-191)

QY	1	ATGACGACACTTCTCTAAACCAAAAGAAAAACCAAAAGAAACCCAA-CCCGGGCCACAG	59
Db	1	MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrLLeArgTrpProL	20
QY	60	GACGTAAAGTTCCACAGGGCGGGTCAGATGTTGGTGGAGTTTACGTGTACACAGCAG	119
Db	21	AspIleLysPheProGlyGlyGlyGlnIleValGlyGlyValTyrValLeuProArgArg	40
QY	120	GGCCCCGAGTGGGTGTGTGCGCGAGTGGCCACACACTCCGACGGTCCCAACTCGCAGT	179
Db	41	GlyProArgLeuGlyValAlaArgAlaThrArgLysLysThrSerGlnArgSerGlnProArgGly	60
QY	180	AGGGCCCAACCACTCCCGAGGGCGCGCCGACCAACCGAGGGCAGTCTCGGCGTCACCGCGG	239
Db	61	ArgArgGlnProIleProLysAlaIleArgArgSerGlnGlyArgSerTrpAlaGlnProGly	80
QY	240	TACCTTGGCCCGCTATATGTGGAAATGAGGGCTCGGGTGGGACAGGGTGCTCTGTCGCCG	299
Db	81	TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro	100
QY	300	CGGCGCTCGCCCGTGTGTGGGGCCAAATGACCCCGCGGCGCAG	344
Db	101	ArgGlySerArgProSerTrpGlyProAsnAspProAlaGlnGlnG	115

```
Search completed: February 19, 2003, 02:15:40
Job time : 47.5 secs
```



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 19, 2003, 02:11:15 ; Search time 22.5 Seconds

(without alignments)
902.303 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 639

Sequence: 1 ATGAGCAGACTCTCTAACC.....AAATGACCCCGCGCAGCA 345

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame-n2p.model -DEV=xlp
-DB=us09_1/us09_1/spool/us09873224/runat_14022003_092706_27514/app.query.fasta_1.519
-LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.col
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09873224.ecgn.1.1.21.6runat_14022003_092706_27514 -NCPU=6 -ICPU=3
-NARN_XLPHY -NO MAP -LARGEDUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-YAPOP=10 -YGAPEXT=0.5 -XGAPEXT=0.5 -YGAPEXT=6 -YGAPEXT=7
-YAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	608	95.1	115	4	US-08-836-075A-50	Sequence 50, App1
2	588	92.0	191	2	US-08-290-665A-187	Sequence 187, App
3	588	92.0	191	2	US-08-290-665A-187	Sequence 187, App
4	588	92.0	191	2	US-08-290-665A-187	Sequence 187, App
5	588	92.0	191	2	US-08-290-665A-187	Sequence 187, App
6	588	92.0	191	2	US-08-290-665A-187	Sequence 187, App
7	588	92.0	191	2	US-08-290-665A-187	Sequence 187, App
8	587	91.9	191	2	US-08-290-665A-189	Sequence 189, App
9	587	91.9	191	2	US-08-290-665A-189	Sequence 189, App
10	574	89.8	191	2	US-08-290-665A-192	Sequence 192, App
11	574	89.8	191	2	US-08-290-665A-192	Sequence 192, App
12	574	89.8	191	2	US-08-290-665A-195	Sequence 195, App

13	574	89.8	191	5	PCT-US95-10398-192	Sequence 192, App
14	574	89.8	191	5	PCT-US95-10398-193	Sequence 193, App
15	574	89.8	191	5	PCT-US95-10398-195	Sequence 195, App
16	571	89.4	191	4	US-08-836-075A-12	Sequence 12, App
17	570	89.2	191	2	US-08-290-665A-196	Sequence 196, App
18	570	89.2	191	2	PCT-US95-10398-196	Sequence 196, App
19	569	89.0	2894	2	US-08-466-975A-23	Sequence 23, App
20	569	89.0	2894	2	US-08-391-671A-23	Sequence 23, App
21	569	89.0	2894	2	US-08-467-902A-23	Sequence 23, App
22	569	89.0	2894	2	US-08-467-902A-23	Sequence 23, App
23	568	88.9	191	2	US-08-290-665A-156	Sequence 156, App
24	568	88.9	191	2	US-08-290-665A-157	Sequence 157, App
25	568	88.9	191	2	US-08-290-665A-158	Sequence 158, App
26	568	88.9	191	2	US-08-290-665A-159	Sequence 159, App
27	568	88.9	191	2	US-08-290-665A-160	Sequence 160, App
28	568	88.9	191	2	US-08-290-665A-161	Sequence 161, App
29	568	88.9	191	2	US-08-290-665A-162	Sequence 162, App
30	568	88.9	191	2	US-08-290-665A-163	Sequence 163, App
31	568	88.9	191	2	US-08-380-160-3	Sequence 160, App
32	568	88.9	191	2	PCT-US95-10398-156	Sequence 156, App
33	568	88.9	191	2	PCT-US95-10398-157	Sequence 157, App
34	568	88.9	191	2	PCT-US95-10398-158	Sequence 158, App
35	568	88.9	191	2	PCT-US95-10398-159	Sequence 159, App
36	568	88.9	191	2	PCT-US95-10398-160	Sequence 160, App
37	568	88.9	191	2	PCT-US95-10398-161	Sequence 161, App
38	568	88.9	191	2	PCT-US95-10398-162	Sequence 162, App
39	568	88.9	191	2	PCT-US95-10398-163	Sequence 163, App
40	568	88.9	191	2	PCT-US95-10398-164	Sequence 164, App
41	568	88.9	191	2	PCT-US95-10398-165	Sequence 165, App
42	568	88.9	191	2	PCT-US95-10398-166	Sequence 166, App
43	568	88.9	191	2	PCT-US95-10398-167	Sequence 167, App
44	568	88.9	191	2	PCT-US95-10398-168	Sequence 168, App
45	568	88.9	191	2	PCT-US95-10398-169	Sequence 169, App

ALIGNMENTS

RESULT 1
US-08-836-075A-50
Sequence 50, Application US/08836075A
Patent No. 6180768
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
APPLICANT: STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
TITLE OF INVENTION: AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:

```

NAME: KAMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-075A-50

Alignment Scores:
Pred. No.: 1,77e-50      Length: 115
Score: 608.00           Matches: 114
Percent Similarity: 99.13%      Conservative: 0
Best Local Similarity: 99.13%    Mismatches: 1
Query Match: 95.15%             Indels: 1
DB: 4                      Gaps: 0

US-09-873-224-147 (1-345) x US-08-836-075A-50 (1-115)

QY 1 ATGAGCACACTTCTTAACCAACCAAGAAACCAACCAACCC-CGGCCACAG 59
    |||||||
Db 1 MetserThreupProlysProginArgLysThrLysArgAsnThrAsn**ArgProgin 20

QY 60 GACGTTAGTCCCGAGCGCGGCGTCAAGATCGTGTGAGATTACGTGCTACACGACG 119
    |||||||
Db 21 AspAllylspheroGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 40

QY 120 GGGCCCCAGTGGGTGTGCGTGCAGTCCGCAACCTCCGACGGGTGCAACCTGCGCAGT 179
    |||||||
Db 41 GlyProginLeuGlyValArgAlaValArgLysThrSerLysSerGlnProArgSer 60

QY 180 AGGCGCCACCAACCCAGGCGCGCCGCAACGAGGCGAGGTCCTGGGCTCAGCCCGG 239
    |||||||
Db 61 ArgArgGlnProLeuProArgAlaArgArgThrGlnLysArgSerTrpAlaGlnProGly 80

QY 240 TACCTTGCCCTATATGGAATGAGGCGTGGCGGTGGGCGAGGTGCTCTGTCCTCCCG 299
    |||||||
Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100

QY 300 CGCGGCTCTCGCCGCTCGTGGGGCCCAATGACCCCGGCGCAGG 344
    |||||||
Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 2
US-08-290-665A-187
Sequence 187, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BURKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994

```

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homomapiens
INDIVIDUAL ISOLATE: HK10
US-08-290-665A-187

Alignment Scores:
Pred. No.: 1,55e-48      Length: 191
Score: 588.00           Matches: 108
Percent Similarity: 96.52%      Conservative: 3
Best Local Similarity: 93.91%    Mismatches: 4
Query Match: 92.02%             Indels: 1
DB: 2                      Gaps: 0

US-09-873-224-147 (1-345) x US-08-290-665A-187 (1-191)

QY 1 ATGAGCACACTTCTTAACCAACCAAGAAACCAACCAACCAACCC-CGGCGCACAG 59
    |||||||
Db 1 MetserThreupProlysProginArgLysThrLysArgAsnThrLysArgArgProgin 20

QY 60 GACGTTAGTCCCGAGCGCGGCGTCAAGATCGTGTGAGATTACGTGCTACACGACG 119
    |||||||
Db 21 AspAllylspheroGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 40

QY 120 GGGCCCCAGTGGGTGTGCGTGCAGTCCGCAACCTCCGACGGGTGCAACCTGCGCAGT 179
    |||||||
Db 41 GlyProArgLeuGlyValArgAlaValArgLysThrSerLysSerGlnProArgSer 60

QY 180 AGGCGCCACCAACCCAGGCGCGCCGCAACGAGGCGAGGTCCTGGGCTCAGCCCGG 239
    |||||||
Db 61 ArgArgGlnProLeuProArgAlaArgArgSerGlnLysArgSerTrpAlaGlnProGly 80

QY 240 TACCTTGCCCTATATGGAATGAGGCGTGGCGGTGGGCGAGGTGCTCTGTCCTCCCG 299
    |||||||
Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100

QY 300 CGCGGCTCTCGCCGCTCGTGGGGCCCAATGACCCCGGCGCAGG 344
    |||||||
Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 3
US-08-290-665A-188
Sequence 188, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BURKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA

```



```

? TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
? TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
? NUMBER OF SEQUENCES: 263
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: MORGAN & FINNEGAN
? STREET: 345 PARK AVENUE
? CITY: NEW YORK
? STATE: NEW YORK
? COUNTRY: USA
? ZIP: 10154
? COMPUTER READABLE FORM:
? MEDIUM TYPE: FLOPPY DISK
? COMPUTER: IBM PC COMPATIBLE
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: WORDPERFECT 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US-08/290,665A
? FILING DATE: 15-AUG-1994
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: RICHARD W. BORK
? REGISTRATION NUMBER: 36,459
? REFERENCE/DOCKET NUMBER: 2026-4116
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 758-4800
? TELEFAX: (212) 751-6849
? TELEX: 421792
? INFORMATION FOR SEO ID NO: 190:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 191 amino acids
? TYPE: amino acid
? STRANDEDNESS: unknown
? TOPOLOGY: unknown
? ORIGINAL SOURCE:
? ORGANISM: homosapiens
? INDIVIDUAL ISOLATE: DK12
US-08-290-665A-190

Alignment Scores:
Pred. No.:      1,55e-48      Length:      191
Score:          588.00        Matches:     108
Percent Similarity: 96.52%    Conservative: 3
Best Local Similarity: 93.91%   Mismatches:  4
Query Match:      92.02%       Indels:      1
DB:               Gaps:       0

US-09-873-224-147 (1-345) x US-08-290-665A-190 (1-191)
QY      1 ATGAGCACACTTCTTAACCAACAAGAAAAACAAAAGAACAACCAA-CGCCGGCCACAG      59
Db      1 MetSerThrLeuProLysProGlnArgIshThrlsArgSstHrllleArgIrproIn      20
QY      60 GAGGTAAAGTTCOCAGGGCGGCGGCACATCGTGTCGTGTGAGATTAGCTGCTACCAACGAGG      115
Db      21 AspValLysPheProIyglYglYlmlYvalGlYglYValItyrValLeuProArGarg      40
QY      120 GGCCCCCAGTTGGATGTCGTCGCGACGTGCGCAAGACTTCCGAGCGGTCCGAACCTCGCAGT      179
Db      41 GlYPrArGluenglyValArgAlathrArGluYshhSerdLuArgSerclnProArggly      60
QY      180 AGCGCCACCATTCGCCAAGGGCGCGCCGCAACCGAGGGCAGGTCTCGGGCTOAGCCCGG      239
Db      61 ArgArgGlnProIleProIylalaArgArGserclnglyArgSerTrPalaglnProgly      80
QY      240 TACCCTTGACCCCTATATAGGAATGAGAGCTGCGGGGCGGAGGGGCTGCTGTCGCCCG      299
Db      81 TyrProTrpProLeuItyglYasnIcnglyCyssgITTPtPalaglYTrpLeuLeuSerPro      100
QY      300 CGCGGCTCTGCGCCGTCGTGGGGGCCCAATGACCCCGCGCAGG      344
Db      101 ArgGlySerArgProSerTrpGlyProAsnaSProArGargarg      115

```

```

PCT-US95-10398-187
; Sequence 187, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290,665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6849
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 187:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: HK10
; PCT-US95-10398-187

Alignment Scores:
Pred. No.: 1.55e-48 Length: 191
Score: 588.00 Matches: 108
Percent Similarity: 96.52% Conservative: 3
Best Local Similarity: 93.91% Mismatches: 4
Query Match: 92.02% Indels: 1
DB: 5 Gaps: 0

US-09-873-224-147 (1-345) x PCT-US95-10398-187 (1-191)
QY 1 ATGAGCAGACTGCTTAACCAAGAAAGAAACCAAGCAACCA-CCCGGCGACAG 59
DB 1 MetSerThrLeuProLysProGlnAlaGlySerThrLysArgAsnThrIleArgArgProGln 20
QY 60 GACGTTAAGTCCACGAGCGGCGGTGATGTTGTGAGATTACGTGCTACACGACGAG 119
DB 21 AspValLysPheProGlyGlyGlnIleValIleGlyValIleValLeuProArgArg 40
QY 120 GCGCCCAAGTGGGTGGTGTGAGTGCAGTCCGACGAGCTCCGACGCGCAACCTGGCAGT 179
DB 41 GlyProArgLeuGlyValAlaArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 60

```

```

QY 180 AGCGGCAACCCATCCCGAGGCGCGGCAACCGAGCGAGTCTCGGCTCAGCCCGG 239
DB 61 ArgArgGlnProIleProLysAlaArgArgSerGluGlyArgSerTrpPalaGlnProGly 80
QY 240 TACCTTGCCCTATATGGAATGAGGCTCGGGTGGGAGGAGGTGCTCTCTCCCG 299
DB 81 TyrProTrpProLeuTyrGlyGlnGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 100
QY 300 CCGCGCTCTCGCCGCTGCGGGGCGGCAATGACCCCGCGGAG 344
DB 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 6
PCT-US95-10398-188
; Sequence 188, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BUKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290,665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 751-6849
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: S52
; PCT-US95-10398-188

Alignment Scores:
Pred. No.: 1.55e-48 Length: 191
Score: 588.00 Matches: 108
Percent Similarity: 96.52% Conservative: 3
Best Local Similarity: 93.91% Mismatches: 4
Query Match: 92.02% Indels: 1
DB: 5 Gaps: 0

```

US-09-873-224-147 (1-345) x PCT-US95-10398-188 (1-191)

QY 1 ATGAGCACACTTCTTAACACCAAGAAACCAAAACACCAACCA-CCCGGCCACAG 59
 |||||||
 Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrLleArgArgProGln 20
 QY 60 GACGTTAAGTTCACAGCGCGGTGACAGTCGTGAGTTACGTACCAACGACAG 119
 |||||||
 Db 21 AspValLysPheProGlyGlyGlnIleValGlyValTyrValLeuProArgArg 40
 QY 120 GCGCCCGCAGTTGGTGTGCGTGCAGAGCTCCGACGCGGTGCAACCTCGCAGT 179
 |||||||
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 60
 QY 180 AGGCGCAACCCATCCCGAGCGCGGTGACAGTCGTGAGTTACGTACCAACGACAG 239
 |||||||
 Db 61 ArgArgGlnProLleProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80
 QY 240 TACCTTGCGCCCTATATGGAATGAGGCGTGGGTGGGAGGAGTCTGCTGTCGCCG 299
 |||||||
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
 QY 300 CGCGGCTCTGCGCGGTGCGGTGCGCAATGACCCCGCGCCAGG 344
 |||||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 7
 PCT-US95-10398-190
 ; Sequence 190, Application PC/TUS9510398
 ; GENERAL INFORMATION:
 ; APPLICANT: BORK, J., MILLER, R. H. AND
 ; APPLICANT: PORCELL, R. H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 263
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/10398
 ; FILING DATE: 15-AUG-1995
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/086,428
 ; FILING DATE: 29 JUNE 1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/290/665
 ; FILING DATE: 15 AUGUST 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: RICHARD W. BORK
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4116
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 190:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 191 amino acids
 ; TYPE: amino acid

STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; ORIGINAL SOURCE:
 ; ORGANISM: homosapiens
 ; INDIVIDUAL ISOLATE: DK12
 PCT-US95-10398-190

Alignment Scores:
 Pred. No.: 1,55e-48 Length: 191
 Score: 588.00 Matches: 108
 Percent Similarity: 96.52% Conservative: 3
 Best Local Similarity: 93.91% Mismatches: 4
 Query Match: 92.02% Indels: 1
 DB: 5 Gaps: 0

US-09-873-224-147 (1-345) x PCT-US95-10398-190 (1-191)

QY 1 ATGAGCACACTTCTTAACACCAAGAAACCAAAACACCAACCA-CCCGGCCACAG 59
 |||||||
 Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrLleArgArgProGln 20
 QY 60 GACGTTAAGTTCACAGCGCGGTGACAGTCGTGAGTTACGTACCAACGACAG 119
 |||||||
 Db 21 AspValLysPheProGlyGlyGlnIleValGlyValTyrValLeuProArgArg 40
 QY 120 GCGCCCGCAGTTGGTGTGCGTGCAGAGCTCCGACGCGGTGCAACCTCGCAGT 179
 |||||||
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 60
 QY 180 AGGCGCAACCCATCCCGAGCGCGGTGACAGTCGTGAGTTACGTACCAACGACAG 239
 |||||||
 Db 61 ArgArgGlnProLleProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80
 QY 240 TACCTTGCGCCCTATATGGAATGAGGCGTGGGTGGGAGGAGTCTGCTGTCGCCG 299
 |||||||
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
 QY 300 CGCGGCTCTGCGCGGTGCGGTGCGCAATGACCCCGCGCCAGG 344
 |||||||
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 8
 US-08-290-665A-189
 ; Sequence 189, Application US/08290665A
 ; Patent No. 5882852
 ; GENERAL INFORMATION:
 ; APPLICANT: BORK, J., MILLER, R. H. AND
 ; APPLICANT: PORCELL, R. H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
 ; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
 ; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
 ; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 263
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/290,665A
 ; FILING DATE: 15-AUG-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: RICHARD W. BORK
 ; REGISTRATION NUMBER: 36,459

;; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
;; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
;; TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
;; NUMBER OF SEQUENCES: 263
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: MORGAN & FINNEGAN
;; STREET: 345 PARK AVENUE
;; CITY: NEW YORK
;; STATE: NEW YORK
;; COUNTRY: USA
;; ZIP: 10154
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY DISK
;; COMPUTER: IBM PC COMPATIBLE
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WORDPERFECT 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/290,665A
;; FILING DATE: 15-AUG-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: RICHARD W. BORK
;; REGISTRATION NUMBER: 36,459
;; REFERENCE/DOCKET NUMBER: 2026-4116
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;;
;; INFORMATION FOR SEQ ID NO: 192:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 191 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; ORIGINAL SOURCE:
;; ORGANISM: homosapiens
;; INDIVIDUAL ISOLATE: 28
;;
US-08-290-665A-192
;;
Alignment Scores:
Pred. No.: 3,366-47 Length: 191
Score: 574.00 Matches: 106
Percent Similarity: 95.65% Conservative: 4
Best Local Similarity: 92.17% Mismatches: 5
Query Match: 89.83% Indels: 1
DB: 2 Gaps: 0
;;
US-09-873-224-147 (1-345) x US-08-290-665A-192 (1-191)
QY 1 ATGAGCACACTTCTTAACCAAGAAACCAAGAAACCAAGCC-CGGCCACAG 59
Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProMet 20
QY 60 GACGTTAAGTCCAGAGCGCGGTGAGATCGTGTGAGATTGACGTCTACCAAGCAGG 119
Db 21 AspValLysPheProGlyGlyGlnLeuValGlyValLysLeuLeuProArgArg 40
QY 120 GGGCCCCAGTTGGGTGTCGTGACAGTCGCAAGACTTCGAGCGGTGCAACCTCGCACT 179
Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
QY 180 AGGCGCCAAACCAACCCAGAGCGCGGTGAGATCGTGTGAGATTGACGTCTACCAAGCAGG 239
Db 61 ArgArgGlnProLeuProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80
QY 240 TACCTTGCGCCCTATATAGGAATGAGGCTGCGGGTGGAGGAGGTGCTCTGCTCCCG 299
Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
QY 300 CCGCGCTCTGCGCGCTGCTGGGGCCCAATGACCCCGCGCCAGG 344
Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 11
US-08-290-665A-193
;; Sequence 193 Application US/08290665A
;; Patent No. 3682852
;;
;; GENERAL INFORMATION:
;; APPLICANT: BORK, J., MILLER, R.H. AND
;; APPLICANT: PURCELL, R.H.
;; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
;; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
;; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
;; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
;; NUMBER OF SEQUENCES: 263
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: MORGAN & FINNEGAN
;; STREET: 345 PARK AVENUE
;; CITY: NEW YORK
;; STATE: NEW YORK
;; COUNTRY: USA
;; ZIP: 10154
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: FLOPPY DISK
;; COMPUTER: IBM PC COMPATIBLE
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WORDPERFECT 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/290,665A
;; FILING DATE: 15-AUG-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: RICHARD W. BORK
;; REGISTRATION NUMBER: 36,459
;; REFERENCE/DOCKET NUMBER: 2026-4116
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;;
;; INFORMATION FOR SEQ ID NO: 193:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 191 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: unknown
;; TOPOLOGY: unknown
;; ORIGINAL SOURCE:
;; ORGANISM: homosapiens
;; INDIVIDUAL ISOLATE: 21
;;
US-08-290-665A-193
;;
Alignment Scores:
Pred. No.: 3,366-47 Length: 191
Score: 574.00 Matches: 106
Percent Similarity: 95.65% Conservative: 4
Best Local Similarity: 92.17% Mismatches: 5
Query Match: 89.83% Indels: 1
DB: 2 Gaps: 0
;;
US-09-873-224-147 (1-345) x US-08-290-665A-193 (1-191)
QY 1 ATGAGCACACTTCTTAACCAAGAAACCAAGAAACCAAGCC-CGGCCACAG 59
Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProMet 20
QY 60 GACGTTAAGTCCAGAGCGCGGTGAGATCGTGTGAGATTGACGTCTACCAAGCAGG 119
Db 21 AspValLysPheProGlyGlyGlnLeuValGlyValLysLeuLeuProArgArg 40
QY 120 GGGCCCCAGTTGGGTGTCGTGACAGTCGCAAGACTTCGAGCGGTGCAACCTCGCACT 179
Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
QY 180 AGGCGCCAAACCAACCCAGAGCGCGGTGAGATCGTGTGAGATTGACGTCTACCAAGCAGG 239
Db 61 ArgArgGlnProLeuProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80

QY 240 TACCTTGCCCTATATGGAATGAGGCTGCGGTCGCGACAGGTGCTCTCTCCCG 299
| | | | |
Db 81 TyrtrotrProleuYrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuSerPro 100
| | | | |
QY 300 CGCGCTCTGCGCCCTGCTGTCGGGCCCAATGACCCCGCGCGCAG 344
| | | | |
Db 101 ArgGlySerArGProSerTrpGlyProAsnAspProArGArG 115
| | | | |
RESULT 12
US-08-290-665A-195
; Sequence 195, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,665A
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 195:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: Z6
; US-08-290-665A-195
Alignment Scores:
Pred. No.: 3.36e-47 Length: 191
Score: 574.00 Matches: 106
Percent Similarity: 95.65% Conservative: 4
Best Local Similarity: 92.17% Mismatches: 5
Query Match: 89.83% Indels: 1
DB: 2 Caps: 0
US-09-873-224-147 (1-345) x US-08-290-665A-195 (1-191)
QY 1 ATGACACACTTCTTAACCAAGAAACCAAGAAACCAACCC-CGCGCACAG 59
| | | | |
Db 1 MetSerTrpAsnProIysProGlnArGlySerThrIysArGAsnTrpAsnArGArGProMet 20
| | | | |
QY 60 GACGTTAGTCCACGAGGCGGCGTGTGTCGAGCTTACGTCACACGACG 119
| | | | |

Db 21 AspValLysPheProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArG 40
| | | | |
QY 120 GCGCCCACTGGTGTCTGTCGCTGCACTGCGCAAGACTTCCAGCGGTGCAACTCGCAGT 179
| | | | |
Db 41 GlyProArGLeuGlyValArGAlaThrArGlySerThrSerGluArGSerGlnProArGly 60
| | | | |
QY 180 AGCGCCCAACCATCCCGCGCGCGCAACCGAGGCGACGCTGTCGAGCTCAGCCCGG 239
| | | | |
Db 61 ArgArGlnProIleProIleProLysAlaArGArGSerGluGlyArGSerTrpAlaGlnProGly 80
| | | | |
QY 240 TACCTTGCCCTATATGGAATGAGGCTGCGGTCGCGACAGGTGCTCTCTCCCG 299
| | | | |
Db 81 TyrtrotrProleuYrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuSerPro 100
| | | | |
QY 300 CGCGCTCTGCGCCCTGCTGTCGGGCCCAATGACCCCGCGCGCAG 344
| | | | |
Db 101 ArgGlySerArGProSerTrpGlyProAsnAspProArGArG 115
| | | | |
RESULT 13
PCT-US95-10398-192
; Sequence 192, Application PC/TUS9510398
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; APPLICANT: PURCELL, R.H.
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10398
; FILING DATE: 15-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/086,428
; FILING DATE: 29 JUNE 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290/665
; FILING DATE: 15 AUGUST 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: RICHARD W. BORK
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 192:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; ORIGINAL SOURCE:
; ORGANISM: homosapiens
; INDIVIDUAL ISOLATE: Z8
; PCT-US95-10398-192
Alignment Scores:
Pred. No.: 3.36e-47 Length: 191

Score: 574.00 Matches: 106
 Percent Similarity: 95.65% Conservative: 4
 Best Local Similarity: 92.17% Mismatches: 5
 Query Match: 89.83% Indels: 1
 DB: 5 Gaps: 0

US-09-873-224-147 (1-345) x PCT-US95-10398-192 (1-191)

QY 1 ATGAGCAGACTTCTTAACCAAGAAAAACCAAAACCAACCC-CGGCCACAG 59
 Db 1 MetserThrasnProLysProGlnArgLysThrLysArgsnThrasnArgArpPromet 20
 QY 60 GACGTTAGTTCACAGCGCGGTCACATCGTTGCTGACGTTTACGTTACACACGAG 119
 Db 21 AspPallysPheProGlnGlyGlyGlnIleValGlyGlyValLeuLeuProArgArg 40
 QY 120 GGGCCCGAGTTGGGTGGCTGCGAGTGGCAAGACTCGGAGGCGGCGCAACCTCCAGT 179
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
 QY 180 AGGCGCCACCCATCCCGAGGCGCGCGCAACCGAGGCGAGTCTCGGCTCAGCCGAG 239
 Db 61 ArgArgGlnProLysProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80
 QY 240 TACCTTGCGCCCTATATGGGAATGAGGCTCGGCGTGGCGAGGCTGCTGCTCCCG 299
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
 QY 300 CGGCGCTCTCGCCGCTGCGTGGGCGCCAAATGACCCCGCGGCGAG 344
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 14
 PCT-US95-10398-193

Sequence 193, Application PC/TUS9510398

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R.H. AND

APPLICANT: PURCELL, R.H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10398

FILING DATE: 15-AUG-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/086,428

FILING DATE: 29 JUNE 1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/290,665

FILING DATE: 15 AUGUST 1994

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 193:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORGANISM: unknown

INDIVIDUAL ISOLATE: Z1

PCT-US95-10398-193

Alignment Scores:

Prod. No.: 3,36e-47

Score: 574.00

Percent Similarity: 95.65%

Best Local Similarity: 92.17%

Query Match: 89.83%

DB: 5 Gaps: 0

US-09-873-224-147 (1-345) x PCT-US95-10398-193 (1-191)

QY 1 ATGAGCAGACTTCTTAACCAAGAAAAACCAAAACCAACCC-CGGCCACAG 59
 Db 1 MetserThrasnProLysProGlnArgLysThrLysArgsnThrasnArgArpPromet 20
 QY 60 GACGTTAGTTCACAGCGCGGTCACATCGTTGCTGACGTTTACGTTACACACGAG 119
 Db 21 AspPallysPheProGlnGlyGlyGlnIleValGlyGlyValLeuLeuProArgArg 40
 QY 120 GGGCCCGAGTTGGGTGGCTGCGAGTGGCAAGACTCGGAGGCGGCGCAACCTCCAGT 179
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
 QY 180 AGGCGCCACCCATCCCGAGGCGCGCGCAACCGAGGCGAGTCTCGGCTCAGCCGAG 239
 Db 61 ArgArgGlnProLysProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80
 QY 240 TACCTTGCGCCCTATATGGGAATGAGGCTCGGCGTGGCGAGGCTGCTGCTCCCG 299
 Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
 QY 300 CGGCGCTCTCGCCGCTGCGTGGGCGCCAAATGACCCCGCGGCGAG 344
 Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 15
 PCT-US95-10398-195

Sequence 195, Application PC/TUS9510398

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R.H. AND

APPLICANT: PURCELL, R.H.

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10398

FILING DATE: 15-AUG-1995

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA: 08/290/665
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4600
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z6
PCT-US95-10398-195

Alignment Scores:
Pred. No.: 3,36e-47 Length: 191
Score: 574.00 Matches: 106
Percent Similarity: 95.65% Conservative: 4
Best Local Similarity: 92.17% Mismatches: 5
Query Match: 89.83% Indels: 1
DB: 5 Gaps: 0

US-09-873-224-147 (1-345) x PCT-US95-10398-195 (1-191)

QY 1 ATGACACACTCTCTAACCACAAAGAAACCAACACACACCC-CGGCCACAG 59
|||||
DB 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgArgPromet 20
60 GACGTTAAGTCCACAGCGCGCGGATCGTTGGTGGAGTTTACGTGCTACCAAGCAGG 119
|||||
DB 21 AspvallLysPheProGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArg 40
QY 120 GCGCCCAAGTGGGTGGGTGCGTGCAGTCCGCAAGACTTCCGAGCGGTCCGCAACTCGCAGT 179
|||||
DB 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
QY 180 AGCGCGCAACCCATCCCGACAGCGCGCGCAACCGAGGCGAGGCTCCTGGGCTCAGCCCGG 239
|||||
DB 61 ArgArgGlnProIleProLysAlaArgArgSerGlnGlyArgSerTrpAlaGlnProGly 80
QY 240 TACCTTGGCCCTATATGGGAATGAGGAGCTGCGGGTGGCGAGGGGTGCTCCTGTCCCG 299
|||||
DB 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTyrProLeuSerPro 100
QY 300 CCGGGCTCTGCGCGCTGCGGGGCCCAATGACCCCGCGCAGG 344
|||||
DB 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

Search completed: February 19, 2003, 02:17:48
Job time : 24.5 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: February 19, 2003, 02:12:10 : Search time 148.5 Seconds
(without alignments)
118.712 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 639
Sequence: 1 ATGACGACACTCTCTAAMC.....AATGACCCCGCGCAGCA 345

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 280518

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame.n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09873224/runat_14022003_092706_27543/app.query.fasta_1.519
-DB=published.Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MATEL=200000000 -USER=US09873224_cgn2_1_1.12_runat_14022003_092706_27543
-NCPU=6 -ICPU=3 -NO_XLIFY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	608	95.1	115	9	US-09-851-138-50
2	608	95.1	115	9	US-09-899-046-148
3	571	89.4	319	9	US-09-851-138-12
4	569	89.0	2894	10	US-09-941-611-23

5	568	88.9	182	9	US-10-104-966-2	Sequence 2, Appl1
6	568	88.9 <td>182</td> <td>10</td> <td>US-09-929-955-2</td> <td>Sequence 2, Appl1</td>	182	10	US-09-929-955-2	Sequence 2, Appl1
7	568	88.9	3011	10	US-09-747-419-20	Sequence 20, Appl1
8	568	88.9	3011	9	US-10-104-966-1	Sequence 1, Appl1
9	568	88.9	3011	9	US-09-891-894-3	Sequence 3, Appl1
10	568	88.9	3011	10	US-09-742-659-4	Sequence 4, Appl1
11	568	88.9	3011	10	US-09-952-572-9	Sequence 9, Appl1
12	568	88.9	3011	10	US-09-929-955-1	Sequence 1, Appl1
13	568	88.9	3012	9	US-09-995-937-2	Sequence 2, Appl1
14	568	88.9	3012	10	US-09-238-076-2	Sequence 2, Appl1
15	567	88.7	117	9	US-09-851-138-28	Sequence 28, Appl1
16	566	88.6	166	9	US-09-899-046-194	Sequence 194, App
17	564	88.3	166	9	US-09-899-046-164	Sequence 164, App
18	564	88.3	3011	10	US-09-995-937-20	Sequence 20, Appl1
19	564	88.3	3011	10	US-09-238-076-20	Sequence 20, Appl1
20	562	87.9	319	9	US-09-851-138-64	Sequence 44, Appl1
21	560	87.6	3011	10	US-09-916-359-2	Sequence 2, Appl1
22	558	87.3	113	10	US-09-921-397-78	Sequence 78, Appl1
23	558	87.3	319	9	US-09-851-138-42	Sequence 42, Appl1
24	558	87.3	809	9	US-09-973-025-50	Sequence 50, Appl1
25	556	87.0	126	9	US-09-899-046-166	Sequence 166, App
26	556	87.0	138	9	US-09-851-138-60	Sequence 60, Appl1
27	553	86.5	158	9	US-09-851-138-66	Sequence 66, Appl1
28	551	86.2	137	9	US-09-851-138-46	Sequence 46, Appl1
29	541	84.7	319	9	US-09-851-138-48	Sequence 48, Appl1
30	539	84.4	166	9	US-09-899-046-152	Sequence 152, App
31	539	84.4	169	9	US-09-899-046-42	Sequence 42, Appl1
32	539	84.4	169	9	US-09-899-046-44	Sequence 44, Appl1
33	539	84.4	319	9	US-09-899-046-50	Sequence 50, Appl1
34	539	84.4	319	9	US-09-899-046-52	Sequence 52, Appl1
35	539	84.4	319	9	US-09-899-046-54	Sequence 54, Appl1
36	538	84.2	319	9	US-09-851-138-18	Sequence 18, Appl1
37	533	83.4	319	9	US-09-899-046-144	Sequence 144, App
38	516	80.8	109	9	US-09-851-138-6	Sequence 6, Appl1
39	506	79.2	139	10	US-09-736-959-17	Sequence 17, Appl1
40	501	78.4	108	9	US-09-851-138-14	Sequence 14, Appl1
41	492	77.0	109	9	US-09-851-138-2	Sequence 2, Appl1
42	470	73.6	318	9	US-09-851-138-76	Sequence 76, Appl1
43	466	72.9	96	9	US-09-899-046-192	Sequence 192, App
44	460	72.0	97	10	US-09-756-875-8	Sequence 8, Appl1
45	426	66.7	103	10	US-09-921-397-77	Sequence 77, Appl1

ALIGNMENTS

RESULT 1
US-09-851-138-50
: Sequence 50, Application US/09851138
: Publication No. US20020183508A1
: GENERAL INFORMATION:
: APPLICANT: MAERTENS, GEERT
: STUYVER, LIEVEN
: TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
: AGENTS
: NUMBER OF SEQUENCES: 207
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ARNOLD, WHITE & DURKEE
: STREET: P.O. BOX 4433
: CITY: HOUSTON
: STATE: TEXAS
: COUNTRY: USA
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Microsoft Word 6.0 / ASCII text output
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/851,138
: FILING DATE: 09-May-2001
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/836,075

```

; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-09-851-138-50

Alignment Scores:
Pred. No.: 1,94e-47 Length: 115
Score: 608.00 Matches: 114
Percent Similarity: 99.13% Conservative: 0
Best Local Similarity: 99.13% Mismatches: 1
Query Match: 95.15% Indels: 1
DB: 9 Gaps: 0

US-09-873-224-147 (1-345) x US-09-851-138-50 (1-115)
QY 1 ATGAGCACACTTCTTAACCAAGAAAACCAAGAACACCAACC-CGGCCACAG 59
    |||||||
Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrAsn***ArgProGln 20
QY 60 GACGTTAAGTTCACGAGCGGCGGTCGATCGTTGCGAGTTTACGTGTACACGACAG 119
    |||||||
Db 61 ArgArgGlnProIleProArgAlaArgArgThrGlnGlyArgSerTrpAlaGlnProGly 80
QY 21 AspValLysPheProGlyGlyGlnIleValGlyGlyValLysValLeuProArgArg 40
Db 21 TyrProTrpProLeuTyrGlyGlnIleValGlyGlyValLysValLeuProArgArg 40
QY 120 GGGCCCAAGTTGGGTGCTGCTGAGTGCAGTGCAGACTTCCGAGCGGTGCAGACCTGCGAGT 179
    |||||||
Db 41 GlyProGlnLeuGlyValArgAlaValArgLysThrSerGlnArgSerGlnProArgSer 60
QY 180 AGGCGCAACCCATCCCGAGGCGCGGCGGACCGAGGAGGAGTCTCGGGCTCAGCCCGG 239
    |||||||
Db 61 ArgArgGlnProIleProArgAlaArgArgThrGlnGlyArgSerTrpAlaGlnProGly 80
QY 240 TACCTTGGCCCCCTATATGGGAATGAGGCTGCGGGTGCAGAGGTGCTCTGTCCCG 299
    |||||||
Db 81 TyrProTrpProLeuTyrGlyGlnIleValGlyGlyValLysValLeuProArgArg 100
QY 300 CGCGCTCTCGCCGCTGCTGCGGCGGCAATGACCCCGCGCAGG 344
    |||||||
Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 2
US-09-899-046-148
; Sequence 148, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 148:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-148

Alignment Scores:
Pred. No.: 1,94e-47 Length: 115
Score: 608.00 Matches: 114
Percent Similarity: 99.13% Conservative: 0
Best Local Similarity: 99.13% Mismatches: 1
Query Match: 95.15% Indels: 1
DB: 9 Gaps: 0

US-09-873-224-147 (1-345) x US-09-899-046-148 (1-115)
QY 1 ATGAGCACACTTCTTAACCAAGAAAACCAAGAACACCAACC-CGGCCACAG 59
    |||||||
Db 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrAsn***ArgProGln 20
QY 60 GACGTTAAGTTCACGAGCGGCGGTCGATCGTTGCGAGTTTACGTGTACACGACAG 119
    |||||||
Db 21 AspValLysPheProGlyGlyGlnIleValGlyGlyValLysValLeuProArgArg 40
QY 120 GGGCCCAAGTTGGGTGCTGCTGAGTGCAGTGCAGACTTCCGAGCGGTGCAGACCTGCGAGT 179
    |||||||
Db 41 GlyProGlnLeuGlyValArgAlaValArgLysThrSerGlnArgSerGlnProArgSer 60
QY 180 AGGCGCAACCCATCCCGAGGCGCGGCGGACCGAGGAGGAGTCTCGGGCTCAGCCCGG 239
    |||||||
Db 61 ArgArgGlnProIleProArgAlaArgArgThrGlnGlyArgSerTrpAlaGlnProGly 80
QY 240 TACCTTGGCCCCCTATATGGGAATGAGGCTGCGGGTGCAGAGGTGCTCTGTCCCG 299
    |||||||
Db 81 TyrProTrpProLeuTyrGlyGlnIleValGlyGlyValLysValLeuProArgArg 100
QY 300 CGCGCTCTCGCCGCTGCTGCGGCGGCAATGACCCCGCGCAGG 344
    |||||||
Db 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 3
US-09-851-138-12
; Sequence 12, Application US/09851138
; Publication No. US20020183508A1
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTI
; AGENTS
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOED, WHITE & DUKKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/851,138
; FILING DATE: 09-May-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,075
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; APPLICATION NUMBER: EP 95870076.7

```

```

1      FILING DATE: 28 Jun 1995
2      ATTORNEY/AGENT INFORMATION:
3          NAME: KAMMERER, PATRICIA A.
4          REGISTRATION NUMBER: 29,775
5          REFERENCE/DOCKET NUMBER: INNS-004
6
7      INFORMATION FOR SEQ ID NO: 12:
8          SEQUENCE CHARACTERISTICS:
9              LENGTH: 319 amino acids
10             TYPE: amino acid
11             TOPOLOGY: linear
12
13      MOLECULE TYPE: peptide
14
15      SEQUENCE DESCRIPTION: SEQ ID NO: 12:
16      US-09-851-138-12
17
18      Alignment Scores:
19      Pred. NO.:          4.8e-44          Length:          319
20      Score:              571.00           Matches:           106
21      Percent Similarity: 94.78%           Conservative:      3
22      Best Local Similarity: 92.17%        Mismatches:       6
23      Query Match:        89.36%           Indels:           1
24      DB:                  9               Gaps:            0
25
26      US-09-873-224-147 (1-345) x US-09-851-138-12 (1-319)
27
28      QY      1  ATGAGCACACTTCCTAAACCAAAAGAAAAACCAAGAAACCAACCC-CGGCCACAG 59
29      Db      1  MetserThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProGln 20
30
31      QY      60  GACGTTAAGTCCCGACGGCGGGGCGGACATCGTGGTCGTGACGTTACGTCGTACCAACGAGG 119
32      Db      21  AspAllyLysPheProGlyGlyGlyGlnIleValAlaGlyValIleuLeuProAlaArg 40
33
34      QY      120  GGCCCCAGTTGGGGGTGTCGCTGCAGTCGACGTCGACAGACTCCGACGGGTGACCACTCGCAGT 179
35      Db      41  GlyProArgLeuGlyValArgAlaIleThrArgLysThrSerGlnArgSerGlnProArgGly 60
36
37      QY      180  AGGCGCCCAACCCATCCCGACGGCGCGCCGCAACCGAGGGCAGGTCTGGGCTCAGCCGGG 239
38      Db      61  ArgArgGlnProIleProLysAlaArgArgProGlnIleArgSerThrPalaGlnProGly 80
39
40      QY      240  TACCCCTTGACCCCTATATGGGAATGAGGGCGTCGGGGTGGGAGGGTGGCTGTCTCCCG 299
41      Db      81  TyrProTrpProLeuTyrAlaAsnGlnGlyCysGlyTyrTrpAlaGlyTrpLeuLeuSerPro 100
42
43      QY      300  CGCGAGCTCTCGCCCGCGCTGGGGGCCCAATATGACCCCGCGCGCAGG 344
44      Db      101  ArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115
45
46      RESULT 4
47      US-09-941-611-23
48      ? Sequence 23, Application US/09941611
49      ? Patent No. US2002010640A1
50      ? GENERAL INFORMATION:
51      ? APPLICANT: DELEYS, ROBERT J
52      ? POLLET, DIRK
53      ? MAERTENS, GEERT
54      ? VAN HEYERSMUN, HUGO
55      ? TITLE OF INVENTION: SYNTHETIC ANTIBODIES FOR THE DETECTION OF
56      ? ANTIBODIES TO HEPATITIS C VIRUS
57      ? NUMBER OF SEQUENCES: 23
58      ? CORRESPONDENCE ADDRESS:
59      ? ADDRESSEE: NIXON & VANDERBYE P.C.
60      ? STREET: 1100 NORTH GLEBE ROAD
61      ? CITY: ARLINGTON
62      ? STATE: VA
63      ? COUNTRY: USA
64      ? ZIP: 22201
65      ? COMPUTER READABLE FORM:
66      ? MEDIUM TYPE: floppy disk
67      ? COMPUTER: IBM PC compatible
68      ? OPERATING SYSTEM: PC-DOS/MS-DOS
69      ? SOFTWARE: PatentIn version #1.0, Version #1.30
70      ? CURRENT APPLICATION DATA:

```

```

? APPLICATION NUMBER: US/09/941,611
? FILING DATE: 30-Aug-2001
? CLASSIFICATION: <Unknown>
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/391,671
? FILING DATE: 1995-02-21
? APPLICATION NUMBER: WO PCT/EP91/02409
? FILING DATE: 13-DEC-1991
? APPLICATION NUMBER: EP 9012424.1.2
? FILING DATE: 14-DEC-1990
? ATTORNEY/AGENT INFORMATION:
? NAME: SADOFF, B.J.
? REGISTRATION NUMBER: 36,663
? REFERENCE/DOCKET NUMBER: 1487-5
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 7038164000
? TELEFAX: 7038164100
? INFORMATION FOR SEQ ID NO: 23:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2894 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Alignment Scores:
Pred. No.:          9,79e+44      Length:      2894
Score:              569.00         Matches:     104
Percent Similarity: 94.78%         Conservative: 5
Best Local Similarity: 90.43%       Mismatches: 1
Query Match:        89.05%         Indels:      0
Db:                 10             Gaps:        0

US-09-873-224-147 (1-345) x US-09-941-611-23 (1-2894)

QY      1 ATGAGCACACTTCTTAACCAACAAGAAAGAAAACCAAAAGAACCAACC-CGGCCACAG 59
       | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB      1 MetserThrIleProLysProGIlnArglysthrIysAlgsanThrAsnArGargProGln 20

QY      60 GACGTGAAGTTCCCAAGCGCGGGGCACATCCTTGTTGAGATTACGTGTACACAGCAGG 119
       | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB      21 AspyallYsPheProGLyglYgInlValJslYglYalTYlrLeuleuProArGarg 40

QY      120 GGCCCCCGATTGGGGTGTCGTGCAGTGCGAAGAAGCTTCGAGCGGTGCAACCTCGCAGT 179
       | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB      41 GlyProArGlLeuGLyAlArGAlArthrArglysthresGluArGserGlnProArGLy 60

QY      180 AGGGCGCAACCCATCCCGAGSGCGCGGCCGAACGAGGAGGCTCTGGGCTCAGCCCCGG 238
       | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB      61 ArgArgGlnPrOlePrOlePrOlyValArGargPrOgLuglYarGlnhrTPraLaGlnPrOgly 80

QY      240 TACCGTTGGCCCATATAGGAATGAGGGGCGGCGGTGGAGGCGAGGTGGCTCTGCCCG 299
       | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB      81 TyrProIrrProLeuTYrgLYasngInelYcslYlrtpaaagLYlrpleuLeuSerPro 100

QY      300 CGCGGCTCTCGCCGCTGCTGGGGGCCCAATAGACCCCGCGGCGCAGG 344
       | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB      101 ArgLyserArGProserTrpGLyprOthrAsPraArGarg 115

RESULT 5
US-10-104-966-2
: Sequence 2, Application US/10104966
: Patent No. US20020155124A1
: GENERAL INFORMATION:
: APPLICANT: Matti Salberg
: APPLICANT: Catharina Huilgren
: TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
: FILE REFERENCE: TRIPEP.23AUSCI

```

```
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus core protein sequence
US-10-104-966-2

Alignment Scores:
Pred. No.: 8,28e-44 Length: 182
Score: 568.00 Matches: 105
Percent Similarity: 94.78% Conservative: 4
Best Local Similarity: 91.30% Mismatches: 6
Query Match: 88.89% Indels: 1
DB: 9 Gaps: 0

US-09-873-224-147 (1-345) x US-10-104-966-2 (1-182)
QY 1 ATGAGCAGACTTCTTAACCAAGAAAACCAAGAAACCAACCAACC-CGGCCACAG 59
Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
QY 60 GACGTTAAGTCCACAGCGCGGTCAGATCGTTGAGATTACCTGACACGACG 119
Db 21 AspValLysPheProGlyGlyGlnIleValIgLylValTyrLeuLeuProArg 40
QY 120 GGGCCCAAGTGGGTGGTGGTGCAGTCCGCAAGACTCCGAGCGGCGAACCCTGCAGT 179
Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerInProArg 60
QY 180 AGCGGCCAACCCATCCCAAGGCGCGCCGAACGAGGAGGTCCTCGGCTCAGCCGG 239
Db 61 ArgArgGlnProIleProLysAlaArgArgProGlnGlyArgThrTrpAlaGlnProGly 80
QY 240 TACCTTGGCCCTATATGGAATGAGGCTGGGGTGGGAGGCTGCTGTCCTCCCG 299
Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTyrPalaglyTyrLeuLeuSerPro 100
QY 300 CGGCGCTCGCCCGCGTGGGGCCCAATGACCCCGCGGCGAGG 344
Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArg 115

RESULT 6
; Sequence 2, Application US/09929955
; Patent No. US20020136740A1
; GENERAL INFORMATION:
; APPLICANT: Matti Salberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23US2
; CURRENT APPLICATION NUMBER: US/09/929,955
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/225,767
; PRIOR FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 182
; TYPE: PRT
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus core protein sequence
US-09-929-955-2

Alignment Scores:
Pred. No.: 8,28e-44 Length: 182
Score: 568.00 Matches: 105
Percent Similarity: 94.78% Conservative: 4
Best Local Similarity: 91.30% Mismatches: 6
Query Match: 88.89% Indels: 1
DB: 10 Gaps: 0

US-09-873-224-147 (1-345) x US-09-929-955-2 (1-182)
QY 1 ATGAGCAGACTTCTTAACCAAGAAAACCAAGAAACCAACCAACC-CGGCCACAG 59
Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
QY 60 GACGTTAAGTCCACAGCGCGGTCAGATCGTTGAGATTACCTGACACGACG 119
Db 21 AspValLysPheProGlyGlyGlnIleValIgLylValTyrLeuLeuProArg 40
QY 120 GGGCCCAAGTGGGTGGTGGTGCAGTCCGCAAGACTCCGAGCGGCGAACCCTGCAGT 179
Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerInProArg 60
QY 180 AGCGGCCAACCCATCCCAAGGCGCGCCGAACGAGGAGGTCCTCGGCTCAGCCGG 239
Db 61 ArgArgGlnProIleProLysAlaArgArgProGlnGlyArgThrTrpAlaGlnProGly 80
QY 240 TACCTTGGCCCTATATGGAATGAGGCTGGGGTGGGAGGCTGCTGTCCTCCCG 299
Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTyrPalaglyTyrLeuLeuSerPro 100
QY 300 CGGCGCTCGCCCGCGTGGGGCCCAATGACCCCGCGGCGAGG 344
Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArg 115

RESULT 7
; Sequence 20, Application US/09747419
; Patent No. US2002015582A1
; GENERAL INFORMATION:
; APPLICANT: Lemon, Stanley
; APPLICANT: Yi, Minkyung
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
; FILE REFERENCE: 265,0007 0101
; CURRENT APPLICATION NUMBER: US/09/747,419
; CURRENT FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/171,909
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Polyprotein
US-09-747-419-20

Alignment Scores:
Pred. No.: 1,21e-43 Length: 3011
Score: 568.00 Matches: 105
Percent Similarity: 94.78% Conservative: 4
Best Local Similarity: 91.30% Mismatches: 6
Query Match: 88.89% Indels: 1
DB: 9 Gaps: 0

US-09-873-224-147 (1-345) x US-09-747-419-20 (1-3011)
QY 1 ATGAGCAGACTTCTTAACCAAGAAAACCAAGAAACCAACCAACC-CGGCCACAG 59
```

```
|||||
Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
QY 60 GAGCTTAAGTTCACGAGCGCGGTGATGATGCTGTTGAGTTCAGTTCACGAGCAG 119
Db 21 AspValLysPheProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArg 40
QY 120 GGGCCCCAGTTGGGTGTCGCGTCAGTGCAGACAGACTTCGAGCGGTGCGACCTCGCAGT 179
Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
QY 180 AGGCGCAACCCATCCCGAGCGCGCGCAACCGAGGCGAGTCTGGCTCAGCCCGG 239
Db 61 ArgArgGlnProIleProLysAlaArgArgProGlnGlyArgThrTrpAlaGlnProGly 80
QY 240 TACCTTGCGCCCTATATGGAATGAGGCTGCGGAGTGGCGAGGTGCTGCTGCTCCCG 299
Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTyrTrpAlaGlyTrpLeuLeuSerPro 100
QY 300 CGCGGCTCTCGCCGCTGCTGCGGCGCAATGACCCCGCGCAGG 344
Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArg 115
```

RESULT 8

```
US-10-104-966-1
; Sequence 1, Application US/10104966
; Patent No. US2002015124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: TRIPEP.23AUSC1
; CURRENT APPLICATION NUMBER: US/10/104,966
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; PRIOR FILING DATE: 2000-08-29
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus sequence
US-10-104-966-1
```

```
Alignment Scores:
Pred. No.: 1,21e-43 Length: 3011
Score: 568.00 Matches: 105
Percent Similarity: 94.78% Conservative: 4
Best Local Similarity: 91.30% Mismatches: 6
Query Match: 88.89% Indels: 1
Db: 9 Gaps: 0
```

US-09-873-224-147 (1-345) x US-10-104-966-1 (1-3011)

```
QY 1 ATGAGCAGACTTCTTAACCAACAAGAAACCAAGAACCAACCC-CGGCCACAG 59
Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
QY 60 GAGCTTAAGTTCACGAGCGCGGTGATGATGCTGTTAGCTTACCTGACACGAGC 119
Db 21 AspValLysPheProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArg 40
QY 120 GGGCCCCAGTTGGGTGTCGCGTCAGTGCAGACAGACTTCGAGCGGTGCGACCTCGCAGT 179
Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
QY 180 AGCGGCAACCCATCCCGAGCGCGCGCAACCGAGGCGAGTCTGGCTCAGCCCGG 239
|||||
```

```
Db 61 ArgArgGlnProIleProLysAlaArgArgProGlnGlyArgThrTrpAlaGlnProGly 80
QY 240 TACCTTGCGCCCTATATGGAATGAGGCTGCGGAGTGGCGAGGTGCTGCTGCTCCCG 299
Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTyrTrpAlaGlyTrpLeuLeuSerPro 100
QY 300 CGCGGCTCTCGCCGCTGCTGCGGCGCAATGACCCCGCGCAGG 344
Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArg 115
```

RESULT 9

```
US-09-891-894-3
; Sequence 3, Application US/09891894
; Publication No. US20030013081A1
; GENERAL INFORMATION:
; APPLICANT: Madson, William
; APPLICANT: Olson, William
; TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITTING HEPATITIS C VIRUS
; FILE REFERENCE: 2048/54896/JFW/SHS
; CURRENT APPLICATION NUMBER: US/09/891,894
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: hepatitis c virus
US-09-891-894-3
```

```
Alignment Scores:
Pred. No.: 1,21e-43 Length: 3011
Score: 568.00 Matches: 105
Percent Similarity: 94.78% Conservative: 4
Best Local Similarity: 91.30% Mismatches: 6
Query Match: 88.89% Indels: 1
Db: 9 Gaps: 0
```

US-09-873-224-147 (1-345) x US-09-891-894-3 (1-3011)

```
QY 1 ATGAGCAGACTTCTTAACCAACAAGAAACCAAGAACCAACCC-CGGCCACAG 59
Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
QY 60 GAGCTTAAGTTCACGAGCGCGGTGATGATGCTGTTGAGTTCAGTTCACGAGCAG 119
Db 21 AspValLysPheProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArg 40
QY 120 GGGCCCCAGTTGGGTGTCGCGTCAGTGCAGACAGACTTCGAGCGGTGCGACCTCGCAGT 179
Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
QY 180 AGGCGCAACCCATCCCGAGCGCGCGCAACCGAGGCGAGTCTGGCTCAGCCCGG 239
Db 61 ArgArgGlnProIleProLysAlaArgArgProGlnGlyArgThrTrpAlaGlnProGly 80
QY 240 TACCTTGCGCCCTATATGGAATGAGGCTGCGGAGTGGCGAGGTGCTGCTGCTCCCG 299
Db 81 TyrProTrpProLeuTyrGlyAsnGlnGlyCysGlyTyrTrpAlaGlyTrpLeuLeuSerPro 100
QY 300 CGCGGCTCTCGCCGCTGCTGCGGCGCAATGACCCCGCGCAGG 344
Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArg 115
```

RESULT 10

```
US-09-742-659-4
; Sequence 4, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Bulkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacquelyn
```

APPLICANT: Lau, Johnson Y.
 APPLICANT: Lemon, Stanley M.
 TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
 FILE REFERENCE: ID01116
 CURRENT APPLICATION NUMBER: US/09/742,659
 CURRENT FILING DATE: 2000-12-21
 PRIOR APPLICATION NUMBER: US 60/171,469
 PRIOR FILING DATE: 1999-12-22
 NUMBER OF SEQ ID NOS: 16
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 4
 LENGTH: 3011
 TYPE: PRT
 ORGANISM: Hepatitis C virus
 US-09-742-659-4

Alignment Scores:

Pred. No.:	1,21e-43	Length:	3011
Score:	568.00	Matches:	105
Percent Similarity:	94.78%	Conservative:	4
Best Local Similarity:	91.30%	Mismatches:	6
Query Match:	88.89%	Indels:	1
DB:	10	Gaps:	0

US-09-873-224-147 (1-345) x US-09-742-659-4 (1-3011)

QY 1 ATGAGCACACTTCTTAACCAAGAAAAACCAAAACCAACCAACC-CGGCCACAG 59
 Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
 QY 60 GACGTTAAGTCCAGGGCGGCGGTGAGATCGTTGGGTTTACGCTCAGCCAGCAGG 119
 Db 21 AspValLysPheProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArg 40
 QY 120 GGGCCCACTGGGTGCGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 179
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
 QY 180 AGCGCCCAACCATCCCAAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 239
 Db 61 ArgArgGlnProIleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGly 80
 QY 240 TACCTTGGCCCTATATGGAATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 299
 Db 81 TyrProTrpProLeuTyrGlyAsnGluGlyGlyGlyTyrPalaGlyTyrPalaLeuSerPro 100
 QY 300 CGCGGCTCTCGCCGCTGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 344
 Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArgArg 115

RESULT 11

US-09-952-572-9
 Sequence 9, Application US/09952572
 Patent No. US20020119495A1
 GENERAL INFORMATION:
 APPLICANT: HAWAII BIOTECHNOLOGY GROUP, Inc.
 APPLICANT: NAKANO, Eileen
 APPLICANT: CLEMENTS, David
 APPLICANT: HUMPHREYS, Tom
 TITLE OF INVENTION: IMMUNOGENIC COMPOSITION OF HEPATITIS C
 FILE REFERENCE: HAWBIO1100
 CURRENT APPLICATION NUMBER: US/09/952,572
 CURRENT FILING DATE: 2001-09-13
 PRIOR APPLICATION NUMBER: US 60/230,927
 PRIOR FILING DATE: 2000-09-13
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 9
 LENGTH: 3011
 TYPE: PRT
 ORGANISM: Hepatitis C Virus
 US-09-952-572-9

Alignment Scores:

Pred. No.:	1,21e-43	Length:	3011
Score:	568.00	Matches:	105
Percent Similarity:	94.78%	Conservative:	4
Best Local Similarity:	91.30%	Mismatches:	6
Query Match:	88.89%	Indels:	1
DB:	10	Gaps:	0

US-09-873-224-147 (1-345) x US-09-952-572-9 (1-3011)

QY 1 ATGAGCACACTTCTTAACCAAGAAAAACCAAAACCAACCAACC-CGGCCACAG 59
 Db 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
 QY 60 GACGTTAAGTCCAGGGCGGCGGTGAGATCGTTGGGTTTACGCTCAGCCAGCAGG 119
 Db 21 AspValLysPheProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArg 40
 QY 120 GGGCCCACTGGGTGCGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 179
 Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly 60
 QY 180 AGCGCCCAACCATCCCAAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 239
 Db 61 ArgArgGlnProIleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGly 80
 QY 240 TACCTTGGCCCTATATGGAATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 299
 Db 81 TyrProTrpProLeuTyrGlyAsnGluGlyGlyGlyTyrPalaGlyTyrPalaLeuSerPro 100
 QY 300 CGCGGCTCTCGCCGCTGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 344
 Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArgArg 115

RESULT 12

US-09-929-955-1
 Sequence 1, Application US/0929955
 Patent No. US20020136740A1
 GENERAL INFORMATION:
 APPLICANT: Matti Sallberg
 APPLICANT: Catharina Hallgren
 TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
 FILE REFERENCE: TRIPEP.23AUS2
 CURRENT APPLICATION NUMBER: US/09/929,955
 CURRENT FILING DATE: 2001-08-15
 PRIOR APPLICATION NUMBER: 09/705,547
 PRIOR FILING DATE: 2000-11-03
 PRIOR APPLICATION NUMBER: 60/229,175
 PRIOR FILING DATE: 2000-08-29
 PRIOR APPLICATION NUMBER: 60/225,767
 PRIOR FILING DATE: 2000-08-17
 NUMBER OF SEQ ID NOS: 49
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 1
 LENGTH: 3011
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Hepatitis C virus sequence
 US-09-929-955-1

Alignment Scores:

Pred. No.:	1,21e-43	Length:	3011
Score:	568.00	Matches:	105
Percent Similarity:	94.78%	Conservative:	4
Best Local Similarity:	91.30%	Mismatches:	6
Query Match:	88.89%	Indels:	1
DB:	10	Gaps:	0

US-09-873-224-147 (1-345) x US-09-929-955-1 (1-3011)

QY 1 ATGAGCACACTTCTTAACCAAGAAAAACCAAAACCAACCAACC-CGGCCACAG 59

```

|||||
Db 1 MetSerThrAsnProLysProGlnArgLysThrIleValArgAsnThrAsnArgArgProGln 20
QY 60 GACGTTAAGTTCACAGGCGCGGTGATGATGCTGTGGAGTTTACCTGCTACCAAGCAGG 119
Db 21 AspValIysPheProGlyGlyGlnIleValGlyValIleuLeuProArgArg 40
QY 120 GGGCCCAAGTGTGGGTGAGTGGCAGACCTTCAGAGGCGGCGCAACCTCGCAGT 179
Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 60
QY 180 AGCGCCCAACCCATCCCAAGGCGCGCCGCAACCGAGGCGAGTCTCGGCTCAGCCGCG 239
Db 61 ArgArgGlnProIleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGly 80
QY 240 TACCTTGAGCCCTATATGGAATGAGGCTGCGGCTGCGAGGCTGCTCTGCTCCCG 299
Db 81 TyrProTrpProLeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
QY 300 CGGCGCTCTCGCCCTGCTGGGCGCCCAATGACCCCGCGCGCAGG 344
Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArgArg 115

RESULT 13
US-09-995-937-2
; Sequence 2, Application US/09995937
; Publication No. US20030028010A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995 937
; FILING DATE: 28-No. US20030028010A1-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3012 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-995-937-2

Alignment Scores:
Pred. No.: 1,21e-43 Length: 3012
Score: 568.00 Matches: 105

```

```

Percent Similarity: 94.78% Conservative: 4
Best Local Similarity: 91.30% Mismatches: 6
Query Match: 88.89% Indels: 1
DB: 9 Gaps: 0

US-09-873-224-147 (1-345) x US-09-995-937-2 (1-3012)
QY 1 ATGACACACTTCCCAAAACCAAAAGAAAACCAAAACACCAACCC-CGCGCACAG 59
Db 1 MetSerThrAsnProLysProGlnArgLysThrIleValArgAsnThrAsnArgArgProGln 20
QY 60 GACGTTAAGTTCACAGGCGCGGTGATGATGCTGTGGAGTTTACCTGCTACCAAGCAGG 119
Db 21 AspValIysPheProGlyGlyGlnIleValGlyValIleuLeuProArgArg 40
QY 120 GGGCCCAAGTGTGGGTGAGTGGCAGACCTTCAGAGGCGGCGCAACCTCGCAGT 179
Db 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 60
QY 180 AGCGCCCAACCCATCCCAAGGCGCGCCGCAACCGAGGCGAGTCTCGGCTCAGCCGCG 239
Db 61 ArgArgGlnProIleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGly 80
QY 240 TACCTTGAGCCCTATATGGAATGAGGCTGCGGCTGCGAGGCTGCTCTGCTCCCG 299
Db 81 TyrProTrpProLeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro 100
QY 300 CGGCGCTCTCGCCCTGCTGGGCGCCCAATGACCCCGCGCGCAGG 344
Db 101 ArgGlySerArgProSerTrpGlyProThrAspProArgArgArg 115

RESULT 14
US-09-238-076-2
; Sequence 2, Application US/09238076
; Patent No. US20020102540A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; REFERENCE/DOCKET NUMBER: 6029-4831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3012 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
US-09-238-076-2

```

Alignment Scores:	
Pred. No.:	1.21e-43
Score:	568.00
Percent Similarity:	94.78%
Best Local Similarity:	91.30%
Query Match:	86.89%
DB:	10
Length:	3012
Matches:	105
Conservative:	4
Mismatches:	6
Indels:	1
Gaps:	0

US-09-873-224-147 (1-345) x US-09-238-076-2 (1-3012)

QY	1	ATGAGCACACTTCTCTAAACCAAAAGAAAAACCAAAAGAAACACCAACCC -CGGACAG	59
Db	1	MetserThrasnProlysProglInargIyThrlsyrAsnThrAsnArgAvcProglIn	20
QY	60	GACCTTAAGTTCCTCCAGGCGGCGGTCAATCGTGTGTGGAGTTTACGTCTACACGACG	119
Db	21	AspAllylsPheProglIyglYglYglInIleValIelYelYalIyTrleuLeuprOAlaArg	40
QY	120	GGCCCCAGTTGGGTGTGCGGTGCAGTGCAGACACTTCCAGGCGTGCACACTCCGAGT	179
Db	41	GlyTrorArgIeuGlYalArgAlaIthrArgIyIthrSerIunArgSerIunProAlaGly	60
QY	180	AGGCGCAACCCAAACCCCAAGGCGGCGGCCAACCAGAGGGCGAGTCTGGCTACGCCGGG	239
Db	61	ArgArgIuinProIleProIySalAlaArgArProglInIyArgIthrPrAlaGlnProGly	80
QY	240	TACCCTTGGCCCCATATATGGGAATGAGGCGTGGCGGTGGGACAGGTCCTCTGTCGCC	299
Db	81	TyrTrorItrProIeuIyrgIyAsnGlugIyCysglYTrpIlaGlYItrPneuIeuSerPro	100
QY	300	CGGCGCTTCGCGCGTGTGGGGGCCCAATAGACCCCGCGCGACG	344
Db	101	ArgIySerIarGProSerItrPglYItrOthAsnProrArgIrrArg	115

```

RESULT 15
US-09-851-138-28
; Sequence 28, Application US/09851138
; Publication NO. US20020183508A1
;
GENERAL INFORMATION:
;
APPLICANT: MAERTENS, GERT
;
STUYVER, LIEVEN
;
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; AGENTS
;
NUMBER OF CLAIMS: 1

```

```

;      REGISTRATION NUMBER: 29,775
;      REFERENCE/DOCKET NUMBER: INNS:004
;      INFORMATION FOR SEQ ID NO: 28:
;      SEQUENCE CHARACTERISTICS:
;          LENGTH: 117 amino acids

```

US-09-851-138-28

Alignment Scores:	
Pred. No.:	9.6e-44
Score:	111
Percent Similarity:	567.00
Best local Similarity:	94.78%
Query Match:	91.30%
DB:	88.73%
	9
Length:	111
Matches:	105
Conservative:	94
Mismatches:	5
Indels:	1
Gaps:	0

US-09-873-224-147 (1-345) x US-09-851-138-28 (1-117)

QY	1	ATATGAGCAGCATTCTCTAAACCCAAAGAAAACCAAAAGAAACACCAACCC -CGGGCCACAG	59
Db	1	MetSerThrAspProLysProGlnArgLysThrLysArgAsnThrAsnArgArgProMet	20
QY	60	GACGTAAAGTCTCCAGGGGGGGGGTCAATGCTGGTGGAGATTACGCTGCACACGAG	119
Db	21	AspValLysPheProGlyLeuLeuIleValGlyValThrLysLeuLeuProArgArg	40
QY	120	GGCCCCCAAGTGGGTGTCGTGCAGTGCAGACATCTCCAGCGGGTGCACCTCGCAGT	179
Db	41	GlyProArgLeuGlyValAlaArgAlaThrArgLysThrSerGlnArgSerGlnProArgGly	60
QY	180	AGGCGCAGCAACCCATCCCGCAGGGCGCGCCAGACGAGCGAGCGCTGGGGTCAGCCGGG	239
Db	61	ArgArgGlnProIleProLysArgArgArgSerGlnGlyArgSerThrPalaGlnProGly	80
QY	240	TACCTTGAGCCCTATATGAGATAGAGGCTCGCGGGTGGGACAGGGTGCCTCTGCCCCG	299
Db	81	TyrProIleProLeuLysArgLysAsnGlnGlyCysGlyTyrPala**TyrPleuLeuSerPro	100
QY	300	CGGGGCTCTGCGCCGCTGCTGGGGGCCCAATGAGACCCCGGGCGAGG	344
Db	101	ArgGlySerArgProSerThrPalaProAsnAspProAlaArgArg	115

Search completed: February 19, 2003, 02:23:02
Job time : 155.5 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 00:50:15 ; Search time 1243 Seconds

(without alignments)
8077.609 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 345
Sequence: 1 ATGAGCACACTTCTTAACC.....AAATGACCCCGCGGCAGGA 345Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4109280Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: gb.vi:*
16: em.ba:*
17: em.fun:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.ov:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pln:*
35: em.htg_rtd:*
36: em.htg_mam:*
37: em.htg_vrt:*
38: em_sy:*
39: em.htgo_hum:*
40: em.htgo_mus:*
41: em.htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	345	100.0	345	6 A40747	A40747 Sequence 14
2	345	100.0	345	6 AX031725	AX031725 Sequence
3	345	100.0	345	6 AX031995	AX031995 Sequence
4	345	100.0	345	6 AX032265	AX032265 Sequence
5	309	89.6	309	6 A50394	A50394 Sequence 49
6	309	89.6	309	6 AR127535	AR127535 Sequence
7	275	79.7	325	6 HC3ND96	X78863 Hepatitis C
8	136	39.4	1584	14 HPCJK030A3	D49747 Hepatitis C
9	94	27.2	1584	14 HPCJK049A5	D49749 Hepatitis C
10	94	27.2	9450	14 HPCJK049E1	D63821 Hepatitis C
11	78	22.6	357	14 HPCCP4	D30047 Hepatitis C
12	64	18.6	1584	14 HPCJK070A8	D49752 Hepatitis C
13	57	16.5	1584	14 HPCJK055A6	D49750 Hepatitis C
14	55	15.9	1584	14 HPCJK072A9	D49753 Hepatitis C
15	47	13.6	342	14 HCU1239	U11239 Hepatitis C
16	47	13.6	342	14 AY089773	AY089773 Hepatitis C
17	47	13.6	366	14 AY089779	AY089779 Hepatitis C
18	47	13.6	9390	14 HCVEN51	X76918 Hepatitis C
19	47	13.6	9425	14 AF046866	AF046866 Hepatitis C
20	44	12.8	168	14 AY079252	AY079252 Hepatitis C
21	44	12.8	342	14 AY089776	AY089776 Hepatitis C
22	44	12.8	360	14 AF482733	AF482733 Hepatitis C
23	44	12.8	366	14 AY089781	AY089781 Hepatitis C
24	44	12.8	1580	14 HPCJK046A4	D49748 Hepatitis C
25	44	12.8	9461	14 HPCJK046E2	D63822 Hepatitis C
26	43	12.5	216	14 AF350918	AF350918 Hepatitis C
27	43	12.5	255	14 AF379062	AF379062 Hepatitis C
28	43	12.5	255	14 AF379097	AF379097 Hepatitis C
29	43	12.5	345	14 HCVCRAN3	D29467 Hepatitis C
30	43	12.5	506	14 HCVHN3	X76410 Hepatitis C
31	43	12.5	549	6 AR097122	AR097122 Sequence
32	43	12.5	549	6 AR130620	AR130620 Sequence
33	43	12.5	549	6 AR171969	AR171969 Sequence
34	43	12.5	573	6 AX030970	AX030970 Sequence
35	43	12.5	573	14 HCU10235	U10235 Hepatitis C
36	43	12.5	831	6 A50410	A50410 Sequence 65
37	43	12.5	831	6 AR127543	AR127543 Sequence
38	43	12.5	1642	14 HPCN51SPK	M74814 Hepatitis C
39	43	12.5	9431	14 HCU45476	D45476 Hepatitis C
40	41	11.9	326	14 HPC1196CB	D43676 Hepatitis C
41	40	11.6	573	14 HCU10237	U10237 Hepatitis C
42	40	11.0	308	14 AF046826	AF046826 Hepatitis C
43	38	11.0	573	14 HCU10210	U10210 Hepatitis C
44	38	11.0	9454	14 HPCK3A	D28917 Hepatitis C
45	36	10.4	527	14 AY070178	AY070178 Hepatitis C

ALIGNMENTS

RESULT 1
A40747
LOCUS A40747 345 bp DNA
DEFINITION Sequence 147 from Patent WO9425601.
ACCESSION A40747
VERSION A40747.1 GI:2296744
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 345)
AUTHORS Maertens,G. and Stuyver,L.
TITLE NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
THERAPEUTIC AND DIAGNOSTIC AGENTS
JOURNAL Patent: WO 9425601-A 147 10-NOV-1994;

COMMENT INNOCENTICS NV (BE)
 Other publication AU 2139100 941110
 Other publication CN 6722294 941121
 Other publication CN 1108030 950906
 Other publication FI 946067 941223
 Other publication NO 944967 941221
 Other publication JP 75084423T 950921.
 Location/Qualifiers

FEATURES
 source
 1..345
 /organism="unidentified"
 /db_xref="taxon:32644"

CDS
 1..>345
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAA02529.1"
 /db_xref="GI:2296745"
 /translation="MSTLPKPKQKRTKNTNPGHRTLSQAANVSLVEFTCYHAGAPSW
 VCVOCARLPSSGRNLAVGANPSPGRAPRAGPGLSPGTLGPMGMRAAGGGGCPRAA
 LARRGAQMTPGAG"

mat_peptide
 1..342
 /product="unnamed"

BASE COUNT 71 a 113 c 110 g 51 t
 ORIGIN

Query Match 100.0%; Score 345; DB 6; Length 345;
 Best Local Similarity 100.0%; Pred. No. 6.4e-193;
 Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCACACTTCTTAACCAAGAAACCAAAAGAAACCAACCAACCCCGCCACAGG 60
 |||
 Db 1 ATGAGCACACTTCTTAACCAAGAAACCAAAAGAAACCAACCAACCCCGCCACAGG 60
 QY 61 ACGTTAAGTTCCAGAGCGGGGTGAGATCGTTGTGAGTTTACGTGCTACACGAGG 120
 |||
 Db 61 ACGTTAAGTTCCAGAGCGGGGTGAGATCGTTGTGAGTTTACGTGCTACACGAGG 120
 QY 121 GCCCCAGTTGGGTGTGAGTGCAGTCCGACACTCCGAGCGGTGCGAACCTCGCAGTA 180
 |||
 Db 121 GCCCCAGTTGGGTGTGAGTGCAGTCCGACACTCCGAGCGGTGCGAACCTCGCAGTA 180
 QY 121 GCCCCAGTTGGGTGTGAGTGCAGTCCGACACTCCGAGCGGTGCGAACCTCGCAGTA 180
 |||
 Db 121 GCCCCAGTTGGGTGTGAGTGCAGTCCGACACTCCGAGCGGTGCGAACCTCGCAGTA 180
 QY 181 GCGGCCAACCCATCCCAAGGCGCGCCGACGAGGAGTCTGAGGCTGACCGGGGT 240
 |||
 Db 181 GCGGCCAACCCATCCCAAGGCGCGCCGACGAGGAGTCTGAGGCTGACCGGGGT 240
 QY 241 ACCCTTGGCCCTATATGGAATGAGGCTGCGGGTGGGCAAGGCTCCTGTCCCGCC 300
 |||
 Db 241 ACCCTTGGCCCTATATGGAATGAGGCTGCGGGTGGGCAAGGCTCCTGTCCCGCC 300
 QY 301 GCGGCTCTCGCCGCTGTGAGGCGCCCAATGACCCCGGCGCAGGA 345
 |||
 Db 301 GCGGCTCTCGCCGCTGTGAGGCGCCCAATGACCCCGGCGCAGGA 345

RESULT 2
 AX031725 345 bp DNA linear PAT 20-SEP-2000
 LOCUS
 DEFINITION Sequence 147 from Patent EP1004670.
 AX031725
 ACCESSION
 VERSION
 KEYWORDS
 ORGANISM
 source
 1..345
 /organism="unidentified"
 /db_xref="taxon:32644"

REFERENCE
 AUTHORS Maertens,G. and Stuyver,L.
 TITLE Sequences of hepatitis C virus genotypes and their use as
 JOURNAL therapeutic and diagnostic agents
 INNOCENTICS NV (BE)
 Patent: EP 1004670-A 147 31-MAY-2000;
 Location/Qualifiers

FEATURES
 source
 1..345
 /organism="unidentified"
 /db_xref="taxon:32644"

CDS
 1..>345
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC09707.1"
 /db_xref="GI:10278925"
 /translation="MSTLPKPKQKRTKNTNPGHRTLSQAANVSLVEFTCYHAGAPSW
 VCVOCARLPSSGRNLAVGANPSPGRAPRAGPGLSPGTLGPMGMRAAGGGGCPRAA
 LARRGAQMTPGAG"

mat_peptide
 1..342
 /product="unnamed"

BASE COUNT 71 a 113 c 110 g 51 t
 ORIGIN

Query Match 100.0%; Score 345; DB 6; Length 345;
 Best Local Similarity 100.0%; Pred. No. 6.4e-193;
 Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCACACTTCTTAACCAAGAAACCAAAAGAAACCAACCAACCCCGCCACAGG 60
 |||
 Db 1 ATGAGCACACTTCTTAACCAAGAAACCAAAAGAAACCAACCAACCCCGCCACAGG 60
 QY 61 ACGTTAAGTTCCAGAGCGGGGTGAGATCGTTGTGAGTTTACGTGCTACACGAGG 120
 |||
 Db 61 ACGTTAAGTTCCAGAGCGGGGTGAGATCGTTGTGAGTTTACGTGCTACACGAGG 120
 QY 121 GCCCCAGTTGGGTGTGAGTGCAGTCCGACACTCCGAGCGGTGCGAACCTCGCAGTA 180
 |||
 Db 121 GCCCCAGTTGGGTGTGAGTGCAGTCCGACACTCCGAGCGGTGCGAACCTCGCAGTA 180
 QY 121 GCCCCAGTTGGGTGTGAGTGCAGTCCGACACTCCGAGCGGTGCGAACCTCGCAGTA 180
 |||
 Db 121 GCCCCAGTTGGGTGTGAGTGCAGTCCGACACTCCGAGCGGTGCGAACCTCGCAGTA 180
 QY 181 GCGGCCAACCCATCCCAAGGCGCGCCGACGAGGAGTCTGAGGCTGACCGGGGT 240
 |||
 Db 181 GCGGCCAACCCATCCCAAGGCGCGCCGACGAGGAGTCTGAGGCTGACCGGGGT 240
 QY 241 ACCCTTGGCCCTATATGGAATGAGGCTGCGGGTGGGCAAGGCTCCTGTCCCGCC 300
 |||
 Db 241 ACCCTTGGCCCTATATGGAATGAGGCTGCGGGTGGGCAAGGCTCCTGTCCCGCC 300
 QY 301 GCGGCTCTCGCCGCTGTGAGGCGCCCAATGACCCCGGCGCAGGA 345
 |||
 Db 301 GCGGCTCTCGCCGCTGTGAGGCGCCCAATGACCCCGGCGCAGGA 345

RESULT 3
 AX031995 345 bp DNA linear PAT 20-SEP-2000
 LOCUS
 DEFINITION Sequence 147 from Patent EP0984068.
 AX031995
 ACCESSION
 VERSION
 KEYWORDS
 ORGANISM
 source
 1..345
 /organism="unidentified"
 /db_xref="taxon:32644"

REFERENCE
 AUTHORS Maertens,G. and Stuyver,L.
 TITLE Sequences of hepatitis C virus genotypes and their use as
 JOURNAL therapeutic and diagnostic agents
 INNOCENTICS NV (BE)
 Patent: EP 0984068-A 147 08-MAR-2000;
 Location/Qualifiers

FEATURES
 source
 1..345
 /organism="unidentified"
 /db_xref="taxon:32644"

CDS
 1..>345
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC09783.1"
 /db_xref="GI:10279108"
 /translation="MSTLPKPKQKRTKNTNPGHRTLSQAANVSLVEFTCYHAGAPSW
 VCVOCARLPSSGRNLAVGANPSPGRAPRAGPGLSPGTLGPMGMRAAGGGGCPRAA
 LARRGAQMTPGAG"

mat_peptide
 1..342
 /product="unnamed"

BASE COUNT 71 a 113 c 110 g 51 t

ORIGIN

Query Match 100.0%; Score 345; DB 6; Length 345;
 Best Local Similarity 100.0%; Pred. No. 6.4e-193;
 Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCACAATCTTCTTAACCAACCAAGAAAAAACCAACCAACCCCGCCACAG 60
 |||||||
 Db 1 ATGAGCACAATCTTCTTAACCAACCAAGAAAAAACCAACCAACCCCGCCACAG 60
 QY 61 ACCTTAAGTTCCAGAGGCGGTGATGCTGTTGAGTTTACGTACCAACGAGGG 120
 |||||||
 Db 61 ACCTTAAGTTCCAGAGGCGGTGATGCTGTTGAGTTTACGTACCAACGAGGG 120
 QY 121 GCCCCAGTTGGGTGTGCGTGCAGTCCGCAAGACTTCCGAGCGGTCCGACCTTCGAGTA 180
 |||||||
 Db 121 GCCCCAGTTGGGTGTGCGTGCAGTCCGCAAGACTTCCGAGCGGTCCGACCTTCGAGTA 180
 QY 181 GGGGCCAACCATCCCGAGGCGCGCCGGAACGAGGCGAGTCTGAGCTCAGCCGGGT 240
 |||||||
 Db 181 GGGGCCAACCATCCCGAGGCGCGCCGGAACGAGGCGAGTCTGAGCTCAGCCGGGT 240
 QY 241 ACCCTTGCCCTATATGGAATGAGGCTGCGGTGAGGAGGTGCTGTCCCGCC 300
 |||||||
 Db 241 ACCCTTGCCCTATATGGAATGAGGCTGCGGTGAGGAGGTGCTGTCCCGCC 300
 QY 301 GCGGCTCTGCGCCCTGCTGAGGCGCCCAATGACCCCGCGCAGGA 345
 |||||||
 Db 301 GCGGCTCTGCGCCCTGCTGAGGCGCCCAATGACCCCGCGCAGGA 345

RESULT 4
 AX032265 345 bp DNA Linear PAT 20-SEP-2000
 LOCUS
 DEFINITION
 AX032265
 ACCESSION
 VERSION
 AX032265.1 GI:10279290
 KEYWORDS
 SOURCE
 ORGANISM
 unclassified.
 unclassified.

REFERENCE
 AUTHORS
 1 (bases 1 to 345)
 Maertens,G. and Stuyver,L.
 Sequences of hepatitis C virus genotypes and their use as
 therapeutic and diagnostic agents
 Patent: EP 0984067-A 147 08-MAR-2000;
 JOURNAL
 INNOGENETICS NV (BE)
 Location/Qualifiers

FEATURES
 source
 1..345
 /organism="unclassified"
 /db_xref="taxon:32644"
 1..>345
 /note="unclassified protein product"
 /codon_start=1
 /protein_id="CAC09860.1"
 /db_xref="gi:10279291"
 /translation="MSTLPKPKRKTNTNPGHRTLLSSQAAVRSIVTEFTCYHAGAPSW
 VCVCARLPKSGRNILAVCANPSPBRAEPBAGPLSPGTLGPYMKHMAAGGSGSPBRAA
 LARGAOMTPGAG"
 1..342
 /product="unnamed"
 BASE COUNT 71 a 113 c 110 g 51 t

Query Match 100.0%; Score 345; DB 6; Length 345;
 Best Local Similarity 100.0%; Pred. No. 6.4e-193;
 Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCACAATCTTCTTAACCAACCAAGAAAAAACCAACCAACCCCGCCACAG 60
 |||||||
 Db 1 ATGAGCACAATCTTCTTAACCAACCAAGAAAAAACCAACCAACCCCGCCACAG 60
 QY 61 ACCTTAAGTTCCAGAGGCGGTGATGCTGTTGAGTTTACGTACCAACGAGGG 120
 |||||||

Db 61 ACCTTAAGTTCCAGAGGCGGTGATGCTGTTGAGTTTACGTACCAACGAGGG 120
 |||||||
 QY 121 GCCCCAGTTGGGTGTGCGTGCAGTCCGCAAGACTTCCGAGCGGTCCGACCTTCGAGTA 180
 |||||||
 Db 121 GCCCCAGTTGGGTGTGCGTGCAGTCCGCAAGACTTCCGAGCGGTCCGACCTTCGAGTA 180
 QY 181 GGGGCCAACCATCCCGAGGCGCGCCGGAACGAGGCGAGTCTGAGCTCAGCCGGGT 240
 |||||||
 Db 181 GGGGCCAACCATCCCGAGGCGCGCCGGAACGAGGCGAGTCTGAGCTCAGCCGGGT 240
 QY 241 ACCCTTGCCCTATATGGAATGAGGCTGCGGTGAGGAGGTGCTGTCCCGCC 300
 |||||||
 Db 241 ACCCTTGCCCTATATGGAATGAGGCTGCGGTGAGGAGGTGCTGTCCCGCC 300
 QY 301 GCGGCTCTGCGCCCTGCTGAGGCGCCCAATGACCCCGCGCAGGA 345
 |||||||
 Db 301 GCGGCTCTGCGCCCTGCTGAGGCGCCCAATGACCCCGCGCAGGA 345

RESULT 5
 A50394 309 bp DNA Linear PAT 07-MAR-1997
 LOCUS
 DEFINITION
 A50394
 ACCESSION
 VERSION
 A50394.1 GI:2303406
 KEYWORDS
 SOURCE
 ORGANISM
 unclassified.
 unclassified.

REFERENCE
 AUTHORS
 1 (bases 1 to 309)
 Maertens,G. and Stuyver,L.
 NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
 PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS
 Patent: WO 9613590-A 49 09-MAY-1996;
 JOURNAL
 INNOGENETICS NV (BE)
 Other publication AU 3844095 960523.
 COMMENT
 FEATURES
 source
 1..309
 /organism="unclassified"
 /db_xref="taxon:32644"
 BASE COUNT 65 a 99 c 97 g 48 t

Query Match 89.6%; Score 309; DB 6; Length 309;
 Best Local Similarity 100.0%; Pred. No. 1.4e-171;
 Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGCACAATCTTCTTAACCAACCAAGAAAAAACCAACCAACCCCGCCACAG 60
 |||||||
 Db 1 ATGAGCACAATCTTCTTAACCAACCAAGAAAAAACCAACCAACCCCGCCACAG 60
 QY 61 ACCTTAAGTTCCAGAGGCGGTGATGCTGTTGAGTTTACGTACCAACGAGGG 120
 |||||||
 Db 61 ACCTTAAGTTCCAGAGGCGGTGATGCTGTTGAGTTTACGTACCAACGAGGG 120
 QY 121 GCCCCAGTTGGGTGTGCGTGCAGTCCGCAAGACTTCCGAGCGGTCCGACCTTCGAGTA 180
 |||||||
 Db 121 GCCCCAGTTGGGTGTGCGTGCAGTCCGCAAGACTTCCGAGCGGTCCGACCTTCGAGTA 180
 QY 181 GGGGCCAACCATCCCGAGGCGCGCCGGAACGAGGCGAGTCTGAGCTCAGCCGGGT 240
 |||||||
 Db 181 GGGGCCAACCATCCCGAGGCGCGCCGGAACGAGGCGAGTCTGAGCTCAGCCGGGT 240
 QY 241 ACCCTTGCCCTATATGGAATGAGGCTGCGGTGAGGAGGTGCTGTCCCGCC 300
 |||||||
 Db 241 ACCCTTGCCCTATATGGAATGAGGCTGCGGTGAGGAGGTGCTGTCCCGCC 300
 QY 301 GCGGCTCTC 309
 |||||||
 Db 301 GCGGCTCTC 309

RESULT 6
ARI27535
LOCUS AF127535 309 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 49 from patent US 6180768.
ACCESSION AF127535
VERSION AF127535.1 GI:14114130
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 309)
AUTHORS Maertens, G., and Stuyver, L.
TITLE Sequences of hepatitis C virus genotypes and their use as
JOURNAL prophylactic, therapeutic and diagnostic agents
FEATURES Patent: US 6180768-A 49 30-JAN-2001;
source Location/Qualifiers
1..309
/organism="unknown"
BASE COUNT 65 a 99 c 97 g 48 t
ORIGIN
Query Match 89.6%; Score 309; DB 6; Length 309;
Best Local Similarity 100.0%; Pred. No. 1,4e-171;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAGCAGACTTCCTTAACCAAGAAAACCAAGAACCAACCCCGCCACAG 60
DB 1 ATGAGCAGACTTCCTTAACCAAGAAAACCAAGAACCAACCCCGCCACAG 60
QY 61 AGCTTAGTCCAGGCGGCGGTGAGATGCTGTGAGTTCAGTTCACACAGG 120
DB 61 AGCTTAGTCCAGGCGGCGGTGAGATGCTGTGAGTTCAGTTCACACAGG 120
QY 121 GCCCCAGTTGGGTGCTGCTGCACTGCGCAAGACTTCAGGCGTCCAGT 180
DB 121 GCCCCAGTTGGGTGCTGCTGCACTGCGCAAGACTTCAGGCGTCCAGT 180
QY 181 GGGCCCAACCCATCCCGAGGCGGCGGCAACGAGGAGTCTCGGCTCAGCC 240
DB 181 GGGCCCAACCCATCCCGAGGCGGCGGCAACGAGGAGTCTCGGCTCAGCC 240
QY 241 ACCCTTGGGCGGCTATATGGAATGAGGGCTGCGGTGGGAGGTGCTCT 300
DB 241 ACCCTTGGGCGGCTATATGGAATGAGGGCTGCGGTGGGAGGTGCTCT 300
QY 301 GCGGCTCTC 309
DB 301 GCGGCTCTC 309
RESULT 7
HC3NL96
LOCUS HC3NL96 325 bp RNA linear VRL 13-JUL-1995
DEFINITION Hepatitis C VIRUS RNA genotype 3 NL96 core.
ACCESSION X78863
VERSION X78863.1 GI:474329
KEYWORDS core region.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus.
REFERENCE 1 (bases 1 to 325)
AUTHORS van Doorn, L.J.
TITLE Direct Submission
JOURNAL Submitted (15-APR-1994) L.J. van Doorn, Diagnostic Centre SSDZ,
Dept. of Molecular Biology, R. de Graafweg 7, PO Box 5010, 2600 GA
Delft, NETHERLANDS
2 (bases 1 to 325)
van Doorn, L.J., Kleter, G.E., Stuyver, L., Maertens, G., Brouwer, J.T.,
Schalm, S.W., Heijltink, R.A., and Quint, W.G.
TITLE Sequence analysis of hepatitis C virus genotypes 1 to 5 reveals
JOURNAL multiple novel subtypes in the Benelux countries
J. Gen. Virol. 76 (Pt 7), 1871-1876 (1995)

MEDLINE 97201609
PUBMED 9049395
FEATURES
source Location/Qualifiers
1..325
/organism="Hepatitis C virus"
/specific_host="Homo sapiens"
/db_xref="taxon:1103"
/cell_line="human plasma"
1..>325
/codon_start=1
/product="core region"
/protein_id="CA55412.1"
/db_xref="GI:474330"
/db_xref="SPTREMBL:O68494"
/translation="MSTLPKPKPKTKRNTNXPQDYKFPFGGQIVGVYVLPKRPOL
GVRAVKTSESRQPSRSRQPIPRARTETRSWAQPEYIPMLPLTGNCGAGWLLSPRG
SRPSWG"
BASE COUNT 65 a 105 c 104 g 50 t 1 others
ORIGIN
Query Match 79.7%; Score 275; DB 14; Length 325;
Best Local Similarity 100.0%; Pred. No. 2,1e-151;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 50 CCGGCACAGAGAGCTTAAGTTCACAGGCGGCGTACATGCTTGAGACTTACGCT 109
DB 51 CCGGCACAGAGAGCTTAAGTTCACAGGCGGCGTACATGCTTGAGACTTACGCT 110
QY 110 ACACAGAGAGGCGCCCGAGTTGGGTGCTGCTGCAAGTTCGAGCGTGC 169
DB 111 ACACAGAGAGGCGCCCGAGTTGGGTGCTGCTGCAAGTTCGAGCGTGC 170
QY 170 ACCTCGAGTAGAGGCGCAACCCATCCCGAGGCGGCGGCAACGAGGCGAGT 229
DB 171 ACCTCGAGTAGAGGCGCAACCCATCCCGAGGCGGCGGCAACGAGGCGAGT 230
QY 230 TCAGCCGCGGATACCTTGGCCCTATATGGAATGAGGGCTGCGGTGGGAGGTGCT 289
DB 231 TCAGCCGCGGATACCTTGGCCCTATATGGAATGAGGGCTGCGGTGGGAGGTGCT 290
QY 290 CCGTCCCCCGGCGGCTGCGCCCGTGTGGGCGC 324
DB 291 CCGTCCCCCGGCGGCTGCGCCCGTGTGGGCGC 325
RESULT 8
HPCJK030A3
LOCUS HPCJK030A3 1584 bp RNA linear VRL 10-FEB-1999
DEFINITION Hepatitis C virus isolate JK030 gene for core, env, and part of
E2/NS1, partial cds.
ACCESSION D49747
VERSION D49747.1 GI:1197102
KEYWORDS core, env, and part of E2/NS1.
SOURCE Hepatitis C virus (isolate:JK030) cDNA to genomic RNA.
ORGANISM Hepatitis C virus.
REFERENCE 1 (sites)
AUTHORS Tokita, H., Okamoto, H., Iizuka, H., Kishimoto, J., Tsuda, F.,
Leshmana, L.A., Miyakawa, Y., and Mayumi, M.
TITLE Hepatitis C virus variants from Jakarta, Indonesia classifiable
JOURNAL into novel genotypes in the second (2e and 2f), tenth (10a) and
MEDLINE eleven (11a) genetic groups
PUBMED 96226020
JOURNAL Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
Immunology Division; Minamikawachi-machi, Kawachi-gun, Tochigi

329-04, Japan (E-mail:hokamoto@jichi.ac.jp,
Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)
Location/Qualifiers

FEATURES

source

1.1584
/organism="Hepatitis C virus"
/isolate="JK030"
/db_xref="taxon:11103"
<1.339
/citation=[2]
/evidence=not_experimental
340.>1584
/citation=[2]
/codon_start=1
/evidence=not_experimental
/product="core, env, and part of E2/NS1"
/protein_id="BAA08581.1"
/db_xref="GI:1197103"

CDS

5'UTR
BASE COUNT 308 a 459 c 453 g 364 t
ORIGIN
Query Match 39.4%; Score 136; DB 14; Length 1584;
Best Local Similarity 99.5%; Pred. No. 6,8e-69;
Matches 186; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 127 AGTTGGGTGTGCTGCAGTCCGACAGACTTCGAGCGGTGCGACCTCCGACAGAGAGCGCC 186
|||||
Db 467 AGTTGGGTGTGCTGCAGTCCGACAGACTTCGAGCGGTGCGACCTCCGACAGAGCGCC 526
|||||
QY 187 AACCCATCCCGAGGCGCGCGACCGAGGCGAGTCTGGGCTGACCGCGGTACCTT 246
|||||
Db 527 AACCCATCCCGAGGCGCGCGACCGAGGCGAGTCTGGGCTGACCGCGGTACCTT 586
|||||
QY 247 GGGCCCTATATGAGGATGAGGCTGGGGTGGGCGAGGTGCTCTGTCCTCCGCGCGGT 306
|||||
Db 587 GGGCCCTATATGAGGATGAGGCTGGGGTGGGCGAGGTGCTCTGTCCTCCGCGCGGT 646
|||||
QY 307 CTCGCC 313
|||||
Db 647 CTCGCC 653
|||||

RESULT 9
HPCJK049A5 1584 bp RNA linear VRL 10-FEB-1999
LOCUS Hepatitis C virus isolate JK049 gene for core, env, and part of
DEFINITION E2/NS1, partial cds.
ACCESSION D49749.1 GI:1197110
VERSION D49749.1
KEYWORDS core, env, and part of E2/NS1.
SOURCE Hepatitis C virus (isolate:JK049) cDNA to genomic RNA.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnaviruses.
REFERENCE 1 (sites)
AUTHORS Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F.,
Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
TITLE Hepatitis C virus variants from Jakarta, Indonesia Classifiable
into novel genotypes in the second (2e and 2f), tenth (10a) and
eleventh (11a) genetic groups
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
JOURNAL 96226020
MEDLINE 2 (bases 1 to 1584)
REFERENCE Okamoto,H.
AUTHORS Unpublished
JOURNAL 3 (bases 1 to 1584)

AUTHORS Okamoto,H.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
Immunology Division: Minamikawachi-machi, Kawachi-gun, Tochigi
329-04, Japan (E-mail:hokamoto@jichi.ac.jp,
Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)
Location/Qualifiers

FEATURES

source

1.1584
/organism="Hepatitis C virus"
/isolate="JK049"
/db_xref="taxon:11103"
<1.339
/citation=[2]
/evidence=not_experimental
340.>1584
/citation=[2]
/codon_start=1
/evidence=not_experimental
/product="core, env, and part of E2/NS1"
/protein_id="BAA08583.1"
/db_xref="GI:1197111"

5'UTR

CDS
BASE COUNT 303 a 461 c 455 g 365 t
ORIGIN

Query Match 27.2%; Score 94; DB 14; Length 1584;
Best Local Similarity 100.0%; Pred. No. 5.5e-44;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 GGTCTGGGCTCAGCGCGGTACCTTGGCCCTATATGAGGATGAGGCTGCGGTGG 279
|||||
Db 560 GGTCTGGGCTCAGCGCGGTACCTTGGCCCTATATGAGGATGAGGCTGCGGTGG 619
|||||
QY 280 CAGGTGCTCTGTCTGCTGCGCGCGGTCTGCGCC 313
|||||
Db 620 CAGGTGCTCTGTCTGCTGCGCGCGGTCTGCGCC 653
|||||

RESULT 10
HPCJK049E1 9450 bp RNA linear VRL 13-FEB-1999
LOCUS Hepatitis C virus isolate JK049 genomic RNA for polyprotein,
DEFINITION complete genome.
ACCESSION D63821.1 GI:1183032
VERSION D63821.1
KEYWORDS polyprotein.
SOURCE Hepatitis C virus (isolate:JK049) cDNA to genomic RNA.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepadnaviruses.
REFERENCE 1 (sites)
AUTHORS Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F.,
Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
TITLE Hepatitis C virus variants from Jakarta, Indonesia Classifiable
into novel genotypes in the second (2e and 2f), tenth (10a) and
eleventh (11a) genetic groups
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
JOURNAL 96226020
MEDLINE 2 (bases 1 to 9450)
REFERENCE Okamoto,H.
AUTHORS Unpublished
JOURNAL 3 (bases 1 to 9450)
REFERENCE Okamoto,H.
AUTHORS Direct Submission
JOURNAL Submitted (10-AUG-1995) Hiroaki Okamoto, Jichi Medical School,
Immunology Division: Minamikawachi-machi, Kawachi-gun, Tochigi

329-04, Japan (E-mail:hokamoto@ichi.ac.jp,
Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)

FEATURES

Source Location/Qualifiers
1..9450
/organism="Hepatitis C virus"
/isolate="JK049"
/db_xref="taxon:11103"
1..9450
/note="genomic RNA for polypeptide precursor, complete genome"
/citation=[2]
/evidence=not_experimental
1..339
/citation=[2]
/evidence=not_experimental
340..9399
/citation=[2]
/codon_start=1
/evidence=not_experimental
/product="polyprotein"
/protein_id="BA09890.1"
/db_xref="GI:1183033"
/translation="MSTLPKPKRTKRNINRRPQDVKFPGGQIVGGVYLPRLPKL
GVRVAKTSEKSRQPRPIPRARTEGRSMADPGYPMPLVGGGGMWMLSPRG
SRPSPMDPRRSRLNGLVIDTLCGRADIMGYIPLVAGVGVGARAALAGVLEAD
GINATGNLPGCSFIFLLALISCLIPATLGRNAGLTVTNDSNGSYEAD
VILHPCITPCVRLNMAKSCWTPYSPYVAPRAATASLETVDVMAATLCSALY
VODLGGALFLVGGGSRHRHRTVODNCSTYGHILGHMADMMNNSPANTLIV
SOVLRLPOTMFDLVIGAHVGVAVAYSMGNAKAFVLCLESVDASTITGVAA
ASGATITSLFSTGAKOPLHVNNGSMHINRTALNCDSINTFGLAYHFNFS
GCVEMSSACSLDRFAGQWGLPRANISGSPSEKPCYHAPRCIDVPAQVCGPY
CETPSVYVATDRKAGAPTYWGENESVFLFESARPTPEWFCCTMMNGSYKTCG
APPCITIKVREGKSNNSLVCPTDCEFRKPDATYRCGAGPWLPRCLVDYRLHNP
CTVNTTIRKVRMGCSGLEHRRNANMTRGRCMLPDRSEMYPLIHSSTEDLILPC
SFVPIALSTGLIHQNTVYVOTIYGISGLVGAIKMEFYILIFLLADARCVL
WMMMLISDAEALENVLUNAISAAGTHGIMWSLVARCVANHVGRFPLAVSVL
WPLLVLMLEPRAYAWMTGDTSTLGAVALFLFPLSPYKIMIRLIMWNTYLA
RCEALQIVPEPLARGARDGILLAGLFPALVEDITKLAILGLYILQASLVV
PYFRAHVVRLCIVRNITGKYVOMVLLALAGFNTYLDHLSPTMDNAEELKL
AAVAPVIESPMEYKIVTGADTAGCDIICGLVSAKICEVILGAPDQVRSQWML
LAPITAHQOTRGLEFVITVSLTGDKNIVYGEIYVLSSTQTLGTSVGVMTYVH
GASRTLAGNKRPALQMTNVDDLVGPPSPGAKSLPCTCGSADLYLITRDVLP
ARRGDSTASLSLSPRLACLGSSSGPIMCSPGVAAGIFRAVCTRGAKALQIPE
SLAQTRSPESDSTPPAVPQTOVGLHAPTSSESTKVPASVAGQVYVLNFS
VAATLGRFEMSHAGIDPNRTGRTITTGAKLITSYGKFLADGGCGGAYVITC
DECHADATSLIGCTVDOAETAGARIVLATAAPPSITVPSNIEVALNCEGEL
PFGRAITPLVTKGGRHLIFCHSKKCKDELAQULSCLVNAVATYRELDSVITQED
VVVCATDALITGCTGDFDPSVIDCNVAEQYVDFSLDPTFSLFETVPODAYSRSQRG
RTGRKSGCTYRVSPGSRSGMFDVSLCEYDAGCAWELTPSETVRLAYLSTEG
LPVCDHLEFMEGVTGLTHIDAHLSQTKOQGNFPLTAYQATCARAALPSPMD
ETWCKLILRPLHGPPLLYRLAGVONEICTTHPVKYIATCMAADEVATSMVLL
GGVMAALAYCLSVGSVYVGHVYIGKGPALVPRKEVYVYOOYDEMECSRAPYITDA
OGIAOFEKVTIGLQOQDOKRADIKPIATYVOKLETFSKHMNYSQIOTAGIS
TLGPNPALASLMAFTASVSTPLTNQTLLENIGGWAASNLAPASTAEVSGSLAA
AVSGIGLKVLDLIDAGAGVAGALVAFKIMGEMPTHEVMAWMLPALISPGALVVG
VJCAALILRRHVPGEGAVOAMNRLIAFASRGNHVAPTHVYESDAKAVYLLSLVY
TOLLRLHOMINEDYTPPCGNMLYDINWVCTYLADEKILKILPKMGNGTGLSC
OKRGRTGRGQVSVTRCGCALLSGHYKNGTMRLYDGRMCANMHGPIGTGTP
STPAPSYASRALMRVADSIVYEVKRVDFRYVITGDDGCKCCQPLPREFTELDG
VRLHRYAVCRRLDDVTFVGLNSVYIGSOLPEPEPDVAVTSMLQDSHITTE
AKRLRDGSPSLASSASOLASPSRATCTTHGRHPDAELITNMLRMDEMGSNITR
VESSEKVVILDSFELPILACDEDELSVAECFKKPKPPALPIAPADVPPLVEEM
KDPVVPPTVAGCALPQKLPVPPPRKRTIVSESVSKALASLEKSPOTCSA
EDESISGVTGOSGSLITPQVOLDDDSDNESHSMPLEGEGRGDDISGSSAGVSGE
OSVCCSMYSWTGALITPCAABEKLIPISLNSLIRHNLVYSTSRSAQOKRY
TFRLQVLDHNTTLKEIKELASGVAKELISVEACLVPSHSARSPFGAGAEVS
LSKALINHSWEDLEDNTPIPIIMANVEFAVAPHKGRKPARLIYPIPLGVR
ICEKRALVDVIOKLPASIMGSAVGYSPKQVRYLKMNSKTPPLGFSYDTRCPS
LPTEDDINVESTIOACDLKDEARVITSLTERLYCGGPMNSKGOHGYRCSAV
LPTSGNTVYCYLAKAKATKAGIKDPSFLVCGDDVYVIAASIDEDKSLRAFTEA
MRTSAPGDDPPPTVDELITSCSNVSAHDAGKRYTLLTRDPEPLARAMEWA
RHTPNSWNLGINIMVPTIWMVIMTHFESILLOAQDEKALDFEMGAYVSTPDL
LPATIERLHGLSAFSLHSYSPVELINRVAGALRKIGIPLPLRAMRRARAVRAKLISQGG

3' UTR

BASE COUNT
ORIGIN

Query Match 27.2% Score 94; DB 14; Length 9450;
Best Local Similarity 100.0%; Pred. No. 6.7e-44;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 GGTCTGGGCTCAGCCGGGTACCTTGCCCTATATGGAATGAGCGCTGCGTGG 279
|||||
Db 560 GGTCTGGGCTCAGCCGGGTACCTTGCCCTATATGGAATGAGCGCTGCGTGG 619
|||||

QY 280 CAGGCTGCTCTGTCCTCCCGCGCGCTGCGCC 313
|||||
Db 620 CAGGCTGCTCTGTCCTCCCGCGCGCTGCGCC 653
|||||

RESULT 11

HPCCP4 357 bp RNA linear VRL 07-FEB-1999
LOCUS
DEFINITION Hepatitis C virus (individual isolate Td-3/93) gene for polypeptide precursor, partial cds (core protein region).
ACCESSION D30047.1 GI:485800
VERSION D30047.1 GI:485800
KEYWORDS core protein.
SOURCE Hepatitis C virus (isolate: Td-3/93) cDNA to genomic RNA.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

REFERENCE

AUTHORS

TITLE

1 (sites)
Hotta, H., Handajani, R., Lusida, M.I., Soemarto, W., Doi, H.,
Miyajima, R. and Homma, M.
Subtype analysis of hepatitis C virus in Indonesia on the basis of
NS5b region sequences
J. Clin. Microbiol. 32 (12), 3049-3051 (1994)

JOURNAL

MEDLINE

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (28-Apr-1994) to DDBJ by:
Hak Hotta
Kobe University School of Medicine
Department of Microbiology
7-5-1 Kusunoki-cho, Chuo-ku
Kobe, Hyogo 650
Japan
Phone: 078-341-7451 x3301
Fax: 078-351-6347

FEATURES

source

CDS

Location/Qualifiers
1..357
/organism="Hepatitis C virus"
/isolate="Td-3/93"
/db_xref="taxon:11103"
1..357
/note="core protein region"
/codon_start=1
/product="polyprotein precursor"
/protein_id="BA06283.1"
/db_xref="GI:485801"
/translation="MSTLPKPKRTKRNINRRPQDVKFPGGQIVGGVYLPRLPKL
GVRVAKTSEKSRQPRPIPRARTEGRSMADPGYPMPLVGGGGMWMLSPRG

BASE COUNT 74 a 117 c 110 g 56 t
ORIGIN

Query Match 22.6%; Score 78; DB 14; Length 357;
Best Local Similarity 100.0%; Pred. No. 1.4e-34;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 GTTGGGTGCTGCTGAGTGGCGAGACTTCCGAGCGGTGCGACCTCGCACTAGCGCGCA 187
|||||
Db 129 GTTGGGTGCTGCTGAGTGGCGAGACTTCCGAGCGGTGCGACCTCGCACTAGCGCGCA 188
|||||
QY 188 ACCCATCCCGAGCGCGC 205
|||||
Db 189 ACCCATCCCGAGCGCGC 206
|||||

RESULT 12

LOCUS HPVJK070A8 1584 bp RNA linear VRL 10-FEB-1999
DEFINITION Hepatitis C virus isolate JK070 gene for core, env, and part of
E2/NS1, partial cds.
ACCESSION D49752.1 GI:1197114
VERSION D49752
KEYWORDS core, env, and part of E2/NS1.
SOURCE Hepatitis C virus (isolate:JK070) cDNA to genomic RNA.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.

REFERENCE

AUTHORS

TITLE

JOURNAL Tokita, H., Okamoto, H., Iizuka, H., Kishimoto, J., Tsuda, F.,
MEDLINE Lesmana, L.A., Miyakawa, Y. and Mayumi, M.
Hepatitis C virus variants from Jakarta, Indonesia classifiable
into novel genotypes in the second (2e and 2f), tenth (10a) and
eleventh (11a) genetic groups
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
96226020

REFERENCE

AUTHORS

TITLE

JOURNAL Okamoto, H.
MEDLINE Unpublished
Hepatitis C virus (bases 1 to 1584)
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
96226020

FEATURES
source
1. 1584
/organism="Hepatitis C virus"
/isolate="JK070"
/db_xref="taxon:11103"
<1. .339
/citation=[2]
/evidence=not_experimental
340. .>1584
/codon_start=1
/citalion=[2]
/product="core, env, and part of E2/NS1"
/protein_id="BA08584.1"
/translation="MSTLPKPOKTKRNTNRPRDVKFPGGQIVGVVYLPKRPRL
GVAVRKTSERSQPRSRQPIPRARTEGRSWAOPGYWPPLYGNEGGAGWLLSPRG
SRPSWGPNDPERRSRNLGKVIDLTGCFADLMGYIPLVGAPVGVARALAHGVRLD
GINFATNLPGCSFSLFLALSLCLPTAGLEVRNAGLYIVTNDGNSISYVACD
IILHPCGVCVVRAGNTSKCWTPISPVAVSRPAATVSLRTYVDMYGAATLCSALY
VGDICGALFLVGOEFSMRHRQMTVOECNCSTYPGHITGHRMAMDMNMSPAVTMY
SSVLRPQTITDLYIGAHGVMAVAYYSMOGNMAKVFVLVCLFSGVDASTTISGSA
ASRWGTSITLFSFGSNOLQIVN"

5'UTR

CDS

BASE COUNT 303 a 460 c 458 g 363 t
ORIGIN

Query Match 18.6%; Score 64; DB 14; Length 1584;
Best Local Similarity 100.0%; Pred. No. 3.4e-26;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 GGNCTGTGGTACCCCGGTACCCTTGGCCCTATATGGAATGAGGCTGCGGTGGG 279
|||||
Db 560 GGNCTGTGGTACCCCGGTACCCTTGGCCCTATATGGAATGAGGCTGCGGTGGG 619
|||||

RESULT 13

LOCUS HPCJK055A6 1584 bp RNA linear VRL 10-FEB-1999
DEFINITION Hepatitis C virus isolate JK055 gene for core, env, and part of
E2/NS1, partial cds.
ACCESSION D49750.1 GI:1197114
VERSION D49750
KEYWORDS core, env, and part of E2/NS1.
SOURCE Hepatitis C virus (isolate:JK055) cDNA to genomic RNA.
ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepacivirus.

REFERENCE

AUTHORS

TITLE

JOURNAL Tokita, H., Okamoto, H., Iizuka, H., Kishimoto, J., Tsuda, F.,
MEDLINE Lesmana, L.A., Miyakawa, Y. and Mayumi, M.
Hepatitis C virus variants from Jakarta, Indonesia classifiable
into novel genotypes in the second (2e and 2f), tenth (10a) and
eleventh (11a) genetic groups
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
96226020

REFERENCE (bases 1 to 1584)
JOURNAL Okamoto, H.
MEDLINE Unpublished
Hepatitis C virus (bases 1 to 1584)
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
96226020

REFERENCE

AUTHORS

TITLE

JOURNAL Okamoto, H.
MEDLINE Unpublished
Hepatitis C virus (bases 1 to 1584)
J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
96226020

FEATURES

source

1. 1584
/organism="Hepatitis C virus"
/isolate="JK055"
/db_xref="taxon:11103"
<1. .339
/citation=[2]
/evidence=not_experimental
340. .>1584
/codon_start=1
/citalion=[2]
/product="core, env, and part of E2/NS1"
/protein_id="BA08584.1"
/translation="MSTLPKPOKTKRNTNRPRDVKFPGGQIVGVVYLPKRPRL
GVAVRKTSERSQPRSRQPIPRARTEGRSWAOPGYWPPLYGNEGGAGWLLSPRG
SRPSWGPNDPERRSRNLGKVIDLTGCFADLMGYIPLVGAPVGVARALAHGVRLD
GINFATNLPGCSFSLFLALSLCLPTAGLEVRNAGLYIVTNDGNSISYVACD
IILHPCGVCVVRAGNTSKCWTPISPVAVSRPAATVSLRTYVDMYGAATLCSALY
VGDICGALFLVGOEFSMRHRQMTVOECNCSTYPGHITGHRMAMDMNMSPAVTMY
SSVLRPQTITDLYIGAHGVMAVAYYSMOGNMAKVFVLVCLFSGVDASTTISGSA
ASRWGTSITLFSFGSNOLQIVN"

5'UTR

CDS

BASE COUNT 311 a 467 c 455 g 351 t
ORIGIN

Query Match 16.5%; Score 57; DB 14; Length 1584;
Best Local Similarity 100.0%; Pred. No. 4.8e-22;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 CCGGCCACAGACGCTTAAGTCCAGCGGGGTCAAGTCGTTGGAGTTACGT 106
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 390 CCGGCCACAGACGCTTAAGTCCAGCGGGGTCAAGTCGTTGGAGTTACGT 446
 RESULT 14
 HPCJK072A9 1584 bp RNA linear VRL 10-FEB-1999
 DEFINITION Hepatitis C virus isolate JK072 gene for core, env, and part of
 E2/NS1, partial cds.
 ACCESSION D49753
 VERSION D49753.1 GI:1197124
 KEYWORDS core, env, and part of E2/NS1.
 SOURCE Hepatitis C virus (isolate:JK072) cDNA to genomic RNA.
 ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 REFERENCE 1 (sites)
 AUTHORS Tokita,H., Okamoto,H., Iizuka,H., Kishimoto,J., Tsuda,F.,
 Lesmana,L.A., Miyakawa,Y. and Mayumi,M.
 TITLE Hepatitis C virus variants from Jakarta, Indonesia classifiable
 into novel genotypes in the second (2e and 2i), tenth (10a) and
 eleventh (11a) genetic groups
 JOURNAL J. Gen. Virol. 77 (Pt 2), 293-301 (1996)
 MEDLINE 96226020
 REFERENCE 2 (bases 1 to 1584)
 AUTHORS Okamoto,H.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1584)
 AUTHORS Okamoto,H.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAR-1995) Hiroaki Okamoto, Jichi Medical School,
 Immunology Division, Minamikawachi-machi, Kawachi-gun, Tochigi
 329-04, Japan (E-mail:hokamoto@jichi.ac.jp,
 Tel:0285-44-2111(ex.3334), Fax:0285-44-1557)
 FEATURES
 source
 1. 1584
 /organism="Hepatitis C virus"
 /isolate="JK072"
 /db_xref="taxon:11103"
 <1..339
 /citation=[2]
 /evidence=not_experimental
 340..>1584
 /citation=[2]
 /codon_start=1
 /evidence=not_experimental
 /product="core, env, and part of E2/NS1"
 /protein_id="BA08587.1"
 /db_xref="GI:1197125"
 /translation="MSTLPKPKRRTKRNTRRPODVKPPGGGQIVGGVYVLPKRGPRPL
 GYRVRKTSERSQPSRROPPIPRARITGRSMADPGVPMPLVNGCGMGLSPG
 SRPSWPNDRPRRSRNIGKVIDTLTCGADLMGTIPLAGAVGVARALAHVRLAD
 GINPAGTMLPGCSFSLPLALISCLPTPLALEYRNSGLVMVNDSSNLSIVEAD
 IILHPGCVPCVSGNTRSRWISISPTVAVPAATASLRTHDMMVGAATLCSALY
 IGDLCALFLVGGGFSRMRHOLMTVOECNCISYPGHLGHRMADMMNMSPATVTV
 SOVLRLPOTIIDIIVIGAHMGVAGVAYVSMGNNAKVFLVLCFLSGVDATHTTTGVA
 ARTTSGLTALTFTTGKHIQILIN"
 BASE COUNT 316 a 465 c 448 g 355 t
 ORIGIN
 Query Match 15.9%; Score 55; DB 14; Length 1584;
 Best Local Similarity 100.0%; Pred. No. 7,4e-21;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 220 GGTCTCGGGCTCAGCCCGGGGTACCTTGCGCCCTATATGGAATAGAGGCTCGG 274
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Db 560 GGTCTCGGGCTCAGCCCGGGGTACCTTGCGCCCTATATGGAATAGAGGCTCGG 614
 RESULT 15
 HCU31239 308 bp RNA linear VRL 05-JUL-2002
 LOCUS

DEFINITION Hepatitis C virus strain EUIND7 isolate Indian core protein (C)
 gene, partial cds.
 ACCESSION U31239
 VERSION U31239.1 GI:973349
 KEYWORDS Hepatitis C virus.
 SOURCE Hepatitis C virus
 ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 REFERENCE 1 (bases 1 to 308)
 AUTHORS Mellor,J., Holmes,E.C., Jarvis,L.M., Yap,P.L. and Simmonds,P.
 TITLE Investigation of the pattern of hepatitis C virus sequence
 diversity in different geographical regions: implications for virus
 classification. The international HCV Collaborative Study Group
 JOURNAL J. Gen. Virol. 76 (Pt 10), 2493-2507 (1995)
 MEDLINE 96030859
 PUBLISHED 7595353
 REFERENCE 2 (bases 1 to 308)
 AUTHORS Holmes,E.C.
 TITLE Direct Submission
 JOURNAL Submitted (07-JUL-1995) Eddie C. Holmes, Zoology, University of
 Oxford, South Parks Road, Oxford OX1 3PS, UK
 FEATURES
 source
 1. 308
 /organism="Hepatitis C virus"
 /strain="EUIND7"
 /isolate="Indian"
 /db_xref="taxon:11103"
 /note="HCV genotype 3"
 <1..>308
 /gene="C"
 <1..>308
 /gene="C"
 /codon_start=1
 /product="core protein"
 /protein_id="AA75053.1"
 /db_xref="GI:973350"
 /translation="KTKRNTIRRPQDVKPPGGGQIVGGVYVLPKRGPRVATKXTS
 ERQPRGRQPIPKARQSEGRSMADPGVPMPLVNGCGMGLSPGSRPSNPNP
 P"
 BASE COUNT 57 a 99 c 99 g 51 t 2 others
 ORIGIN
 Query Match 13.6%; Score 47; DB 14; Length 308;
 Best Local Similarity 100.0%; Pred. No. 3,5e-16;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 264 GAGGCTGCGGGGTGCGCAGGCTGCTCTGTCCCGCGGCGCTCTCG 310
 ||||||||||||||||||||||||||||||||||||||||||||||||
 Db 238 GAGGCTGCGGGGTGCGCAGGCTGCTCTGTCCCGCGGCGCTCTCG 284
 Search completed: February 19, 2003, 01:43:04
 Job time : 1247 secs


```

XX Maertens G, Stuyver L;
XX WPI: 1994-358277/44.
XX P-PSDB; AAR63551.
XX
XX New polynucleotide sequences from hepatitis C virus - and related
XX vectors, polypeptide(s) and antibodies, useful for immunisation,
XX treatment, diagnosis and typing of HCV isolates
XX
XX Claim 2; Page 197-198; 404pp; English.
XX
XX Compositions comprising at least 5, and pref. 8 or more contiguous
XX nucleotides selected from an HCV type 3 genomic sequence, more
XX particularly (i) the region spanning positions 417-957 of the
XX Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
XX 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
XX positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
XX region spanning positions 8023-8235 of the NS5 region of the
XX subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
XX sequence, may be used as primers to amplify nucleic acid from an
XX isolate belonging to a specific genotype, or as a probe for specific
XX detection/classification of nucleic acid. Polypeptides encoded by
XX the nucleotides in such compositions may be used for immunisation
XX against HCV, for the detection of antibodies directed against HCV
XX and for serotyping. This sequence corresponds to the core region
XX of HCV.
XX
SQ Sequence 345 BP; 71 A; 113 C; 110 G; 51 T; 0 other;
Query Match 100.0%; Score 345; DB 15; Length 345;
Best Local Similarity 100.0%; Pred. No. 8.1e-165;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGAGCAGACTTCTTAAACCAAGAAAGAAACCAAGAACACACCCCGGCACAG 60
DB 1 ATGAGCAGACTTCTTAAACCAAGAAAGAAAGAAACCAAGAACACACCCCGGCACAG 60
OY 61 ACCTTAAGTCCAGCGCGGTCAGATCGTGTGAGTTACGTCTACACGACGAGG 120
DB 61 ACCTTAAGTCCAGCGCGGTCAGATCGTGTGAGTTACGTCTACACGACGAGG 120
OY 121 GCCCCAGTGGGTGGTGGTGCAGTGGCGAAGACTTCCAGCGGTGCAACCTGCAGTA 180
DB 121 GCCCCAGTGGGTGGTGGTGCAGTGGCGAAGACTTCCAGCGGTGCAACCTGCAGTA 180
OY 181 GCGCGCAACCCATCCAGGCGCGCGAAGCCGAGGCGAGTCTGCGCTGCAAGTA 180
DB 181 GCGCGCAACCCATCCAGGCGCGCGAAGCCGAGGCGAGTCTGCGCTGCAAGTA 180
OY 241 ACCCTTGCGCCCTATATGGAATGAGGCTGCGGTGGCGAGGCTGCTGCTGCTGCTG 240
DB 241 ACCCTTGCGCCCTATATGGAATGAGGCTGCGGTGGCGAGGCTGCTGCTGCTGCTG 240
OY 301 GCGGCTCTGCGCGGTGCGGTGGCGCAAAATGACCCCGCGGAGGA 345
DB 301 GCGGCTCTGCGCGGTGCGGTGGCGCAAAATGACCCCGCGGAGGA 345
RESULT 2
AAT27961
ID AAT27961 standard; DNA; 310 BP.
XX
XX AAT27961;
XX
XX 11-MAR-1997 (first entry)
XX
XX Hepatitis C virus type 10a isolate NM98 bases 1-310.
XX
XX Hepatitis C virus; subtype; polymerase chain reaction; amplification;
XX PCR; primer; probe; antibody; infection; ss.
XX
XX Hepatitis C virus.

```

```

XX Key Location/Qualifiers
XX Key 50
XX unsure
XX
XX W09613590-A2.
XX
XX 09-MAY-1996.
XX
XX 23-OCT-1995; 95MO-EP04155.
XX
XX 28-JUN-1995; 95EP-0870076.
XX
XX 21-OCT-1994; 94EP-0870166.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Maertens G, Stuyver L;
XX WPI: 1996-251460/25.
XX P-PSDB; AAR96550.
XX
XX Hepatitis C virus polynucleotide unique to unidentified sub: type
XX - used to develop probes and primers for new sub: types and vaccines
XX to prevent and treat infection
XX
XX Claim 6; Fig 3; 150pp; English.
XX
XX The sequences AAT27937-T27989 represent novel sequences isolated from
XX hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
XX 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
XX untranslated region (UR), the Core/E1, NS4 or NS5B regions of the
XX genome. This sequence represents nucleotides 1-310 from the HCV type 10a
XX isolate NE98.
XX
XX The new HCV types were isolated from patients with chronic HCV from the
XX CC Benelux countries, France, Cameroon and Vietnam, because of their
XX aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
XX amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
XX sequenced either directly or partially and used to classify the new
XX CC viruses into (sub)types based on comparison with known sequences.
XX CC The sequences were used to generate the peptides AAR96424-R96524. The
XX CC sequences can also be used to synthesise probes and primers for the
XX CC detection of HCV in a sample. The polypeptides can be used to detect
XX CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
XX
XX Sequence 310 BP; 65 A; 99 C; 97 G; 48 T; 1 other;
Query Match 75.4%; Score 260; DB 17; Length 310;
Best Local Similarity 100.0%; Pred. No. 9.2e-122;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 50 CCGGCGACAGAGCTTAACTCCAGGCGCGGTGAGATCTGTGAGATTACGTGCT 109
DB 51 CCGGCGACAGAGCTTAACTCCAGGCGCGGTGAGATCTGTGAGATTACGTGCT 109
OY 110 ACCAGCAGAGGCGCCCAAGTGGGTGCTGCGTGCAGTGCAGAACTTCCAGCGGTGCT 110
DB 111 ACCAGCAGAGGCGCCCAAGTGGGTGCTGCGTGCAGTGCAGAACTTCCAGCGGTGCT 110
OY 170 ACCCTGAGTAGAGCGCCAGCAACCATCCCAAGGCGCGCGAAGCGAGGAGTCTGCGGC 170
DB 171 ACCCTGAGTAGAGCGCCAGCAACCATCCCAAGGCGCGCGAAGCGAGGAGTCTGCGGC 229
OY 230 TCAGCCCGGTAACCTTGGCCCTATATGGAATGAGGCTGCGGTGGCGAGGCTGCT 230
DB 231 TCAGCCCGGTAACCTTGGCCCTATATGGAATGAGGCTGCGGTGGCGAGGCTGCT 289
OY 290 CCGTCCCGCGCGGCTGCT 309
DB 291 CCGTCCCGCGCGGCTGCT 310

```


XX (INNO-) INNOGENETICS NV.
 PA Maertens G, Stuyver L;
 XX
 PI
 DR WPI: 1996-251460/25.
 DR P-PSDB; AAR96538.
 XX
 PT Hepatitis C virus poly:nucleic acid unique to unidentified subtype
 PT - used to develop probes and primers for new subtypes and vaccines
 PT to prevent and treat infection
 PS
 XX
 PS Claim 6; Fig 3; 150pp; English.
 CC The sequences AAT27937-127989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-1, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5',
 CC untranslated region (UTR), the Core/EL, NS4 or NS5B regions of the
 CC genome. This sequence represents nucleotides (-15)-816 from the HCV type
 CC 1g isolate FR16.
 CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UTR, Core/EL and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
 CC
 SQ Sequence 831 BP; 150 A; 263 C; 244 G; 172 T; 2 other:
 Query Match 12.5%; Score 43; DB 17; Length 831;
 Best Local Similarity 100.0%; Pred. No. 7.3e-12;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 211 CCGAGGCGAGGTCTGGGCTCAGCCGGGTACCTTGCCCTT 253
 Db 227 CCGAGGCGAGGTCTGGGCTCAGCCGGGTACCTTGCCCTT 269
 RESULT 6
 AAT16649
 ID AAT16649 standard; cDNA; 573 BP.
 AC AAT16649;
 XX
 DT 01-OCT-1996 (first entry)
 DE Hepatitis C virus isolate 25 core protein gene.
 XX
 KM HCV; EL; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KM hepatitis; ss.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..573
 FT /tag= a
 FT /product= core_protein
 FT /note= "does not contain stop codon"
 XX
 PN W09605315-A2.
 XX
 PD 22-FEB-1996.
 PD
 XX 15-AUG-1995; 95WO-US10398.
 PF
 XX 15-AUG-1994; 94US-0290665.
 PR
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA

PA (USSH) US SEC DEPT HEALTH.
 XX
 XX Bukh J, Miller RH, Purcell RH;
 PI
 DR WPI: 1996-139709/14.
 DR P-PSDB; AAR92975.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection
 PT
 PS
 XX
 PS Claim 3; Page 173; 340pp; English.
 CC AAT16610-116661 are cDNAs encoding a core protein gene of 52 HCV
 CC isolates. The isolated sequences are useful for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
 CC used in the prevention of HCV infection.
 CC
 SQ Sequence 573 BP; 103 A; 175 C; 173 G; 122 T; 0 other:
 Query Match 11.6%; Score 40; DB 17; Length 573;
 Best Local Similarity 100.0%; Pred. No. 2.5e-10;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 211 CCGAGGCGAGGTCTGGGCTCAGCCGGGTACCTTGCC 250
 Db 212 CCGAGGCGAGGTCTGGGCTCAGCCGGGTACCTTGCC 251
 RESULT 7
 AAT16643
 ID AAT16643 standard; cDNA; 573 BP.
 AC AAT16643;
 XX
 DT 01-OCT-1996 (first entry)
 DE Hepatitis C virus isolate S52 core protein gene.
 XX
 KM HCV; EL; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KM hepatitis; ss.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..573
 FT /tag= a
 FT /product= core_protein
 FT /note= "does not contain stop codon"
 XX
 PN W09605315-A2.
 XX
 PD 22-FEB-1996.
 PD
 XX 15-AUG-1995; 95WO-US10398.
 PF
 XX 15-AUG-1994; 94US-0290665.
 PR
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.
 XX
 PI Bukh J, Miller RH, Purcell RH;
 PI
 DR WPI: 1996-139709/14.
 DR P-PSDB; AAR92969.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection
 PT
 PS Claim 3; Page 169; 340pp; English.

Sequence 573 BP; 111 A; 178 C; 166 G; 118 T; 0 other;

RESULT 1C

CC This sequence is derived from the Hepatitis C virus (HCV)
 CC non-structural protein 5 using primers based on the 5' noncoding
 CC region. The sequence was found to be highly conserved with the 5'
 CC terminus of the cDNA of other HCV-RNAs although it showed slight
 CC variation.. On the basis of this variation it was possible to
 CC identify regions which could be used as probes and primers (See
 CC AA070441-70444) for HCV and were capable of determining HCV genotypes.
 CC
 CC Sequence 803 BP; 162 A; 240 C; 254 G; 317 T

Query Match	10.18;	Score 35;	DB 15;	Length 803;
Best Local Similarity	100.0%;	Pred. No. 8.3e-08;		
Matches	35;	Conservative 0;	Mismatches 0;	

QY	Days	0Y
264		GAGGGCTGCGGGTGGGCAAGGTGGCTCTGCCCC 298
562		GAGGGCTGCGGGTGGGCAAGGTGGCTCTGCCCC 596
Db		

	RESULT	13
ID	AAT16642	
XX	AAT16642 standard; cDNA; 573 BP.	

Oct-1996 (first entry)

hepatitis C virus isolate HK10 core protein gene.

hepatitis; ss.

NEPALALIS C VIRUS.

	Location/Qualifiers
CDS	1..573
T	

```

/note="does not contain stop codon"

```

W09605315-A2

22-FEB-1996.

TO HQS 1993; 95WO-US10398.

4-2600 1994; 94US-0290665.

(USSH) US SEC DEPT HEALTH.

BURN J, MILLER RH, PURCELL RH;

WFL: 1996-139709/14
P-PSDB: AAB92968

F-PSDB; AAR929668

DNA and amino acid sequence of HCV envelope 1 and 2 used to determine the HCV genotype.

Volume 3; Page 169; 340pp; English.

CC isolated. The isolated sequences of a core protein gene of 52 HCV
CC useful for detecting the presence of HCV in the prod. of primers
CC can be used for HCV genotyping. Proteins encoded by the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC proteins may also be used to detect antibodies against HCV infection. The
CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
CC used in the prevention of HCV infection.

XX
SQ Sequence 573 BP; 111 A; 179 C; 166 G; 112 T; 0 N.

Query Match	9.9%;	Score 34;	DB 17;	Length 573
Best Local Similarity	100.0%;	Pred. No. 2.7e-07;		
Matches 34;	Conservative 0;	Mismatches 0		

OY	220	GATCTGGGCTCA6CCCGGGTACCTTGAGCCCT	253
Db	221	GATCTGGGCTCA6CCCGGGTACCTTGAGCCCT	254

RESULT 14
AAQ70439
ID AAQ70439 standard; cDNA to mRNA; 803 BP.
XX

1990 (first entry)

Hepatitis C virus: how used in detection of HCV.

non-structural protein; ss.

1. CONFIDENTIAL (US-114).

Y- T t T O T T z w

20-MAY-1994

20 NOV-1993; 93CA-2110141.

320P-0354370.

(IMMO) IMMUNO JAPAN KK.

Okamoto H.,

WFL; 1994-249721/31.

detecting HCV and establishing its genotype

32pp; English

non-structural protein 5 using primers based on the 5' noncoding region. The sequence was found to be highly conserved with the 5' terminus of the cDNA of other HCV-RNAs although it showed slight variation. On the basis of this variation it was possible to identify regions which could be used as probes and primers (See AA070441-70444) for HCV and were capable of determining

271 C; 234 G; 148 T; 0 other,

Query Match	9.98;	Score 34;	DB 15;	Length 803
Best Local Similarity	100.0%;	Pred. No. 2.7e-07;		
Matches 34;	Conservative	0;	Mismatches	

```

220 GGTCTCTGGGCTCA GCCCGGGTACCTTGGGCCCT 253
    |||||
518 GGTCTTGGGCTCA GCCCGGGTACCTTGGGCCCT 551

```

JUL 15

AAQ70100 standard; cDNA to mRNA; 803 BP

18-MAY-1995 (first entry)

Thu Feb 20 14:41:26 2003

us-09-873-224-147.oli.rng

```

XX      Recombinant oligonucleotide Mlt/92 used in detection of HCV.
DE      Hepatitis C virus; HCV; genotype; detection; identification;
XX      diagnosis; primer; probe; non-structural protein; ss.
KW      Hepatitis C virus (Mlt/92).
XX      CA2110141-A.
XX      28-MAY-1994.
XX      26-NOV-1993; 93CA-2110141.
XX      27-NOV-1992; 92JP-0354370.
XX      (IMMO ) IMMUNO JAPAN INC.
XX      (IMMO ) IMMUNO JAPAN KK.
XX      Okamoto H;
XX      WPI; 1994-249721/31.
XX      New recombinant nucleic acid from hepatitis C virus - useful for
XX      detecting HCV and establishing its genotype
XX      Claim 1; : 32pp; English.
XX      This sequence is derived from the Hepatitis C virus (HCV)
XX      non-structural protein 5 using primers based on the 5' noncoding
XX      region. The sequence was found to be highly conserved with the 5'
XX      terminus of the cDNA of other HCV-RNAs although it showed slight
XX      variation. On the basis of this variation it was possible to
XX      identify regions which could be used as probes and primers (See
XX      AA070441-70444) for HCV and were capable of determining HCV genotypes.
XX      Sequence 803 BP; 163 A; 244 C; 249 G; 147 T; 0 other;
SQ      Query Match          9.9%; Score 34; DB 15; Length 803;
      Best Local Similarity 100.0%; Pred. No. 2.7e-07;
      Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      220 GGTCTGGGCTCAGCCGGGTACCTTGCCCT 253
      |||||||
DB      518 GGTCTGGGCTCAGCCGGGTACCTTGCCCT 551

```

Search Completed: February 19, 2003, 01:22:08
 Job time : 227 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 01:14:55 : Search time 1460 seconds
(without alignments)
3827.016 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 345
Sequence: 1 ATGACACACACTTCCTTAAC.....AATGACCCCGCGCAGGA 345

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estda:*
2: em_estdm:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: qb_est1:*
10: qb_est2:*
11: qb_hic:*
12: qb_est3:*
13: qb_est4:*
14: qb_est5:*
15: em_estfun:*
16: em_estom:*
17: qb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	5.8	151	9	AT605049
2	20	5.8	315	12	BF456065
3	20	5.8	457	10	BB697776
4	20	5.8	576	10	AM563834
5	20	5.8	509	12	BG310307
6	20	5.8	925	12	BF179413

7	19	5.5	222	9	AA853727	AA853727 NHTBCae07
8	19	5.5	240	17	AZ239458	AZ239458 RPT-23-7
9	19	5.5	338	14	D79948	D79948 HMM362E05B
10	19	5.5	402	13	BI307759	BI307759 TYID03 he
11	19	5.5	457	10	BB679924	BB679924 BB679924
12	19	5.5	469	10	AM200309	AM200309 da1e02.y
13	19	5.5	471	14	BO789910	BO789910 hage004ab
14	19	5.5	524	12	BG569327	BG569327 602588634
15	19	5.5	540	17	AC799264	AC799264 HS_5451.B
16	19	5.5	559	9	AJ42822	AJ42822 AJ42822
17	19	5.5	562	17	AZ462011	AZ462011 1M0269C11
18	19	5.5	566	17	B52047	B52047 CIT-HSP-386
19	19	5.5	567	17	CNS0431A	AL272215 Tetradon
20	19	5.5	619	10	AV970225	AV970225 AV970225
21	19	5.5	635	10	BE235455	BE235455 602025261
22	19	5.5	637	10	BE272795	BE272795 601104860
23	19	5.5	676	17	AZ985241	AZ985241 2M0267102
24	19	5.5	677	13	BI267794	BI267794 NF15H121
25	19	5.5	687	13	BI393812	BI393812 ppp1n.pk0
26	19	5.5	697	13	BI393236	BI393236 ppp1n.pk0
27	19	5.5	713	10	AV973044	AV973044 AV973044
28	19	5.5	727	17	AG011398	AG011398 Homo sapi
29	19	5.5	728	17	AG011399	AG011399 Homo sapi
30	19	5.5	768	17	AG171625	AG171625 Pan. trogl
31	19	5.5	783	12	BG616849	BG616849 602615993
32	19	5.5	827	14	BM779785	BM779785 EST590361
33	19	5.5	872	12	BG563095	BG563095 602581887
34	19	5.5	940	17	AZ675194	AZ675194 ENIK92TF
35	19	5.5	1007	13	BI488777	BI488777 603021362
36	19	5.5	1026	14	BO884995	BO884995 ACBNCOURT
37	19	5.5	1138	14	BO877217	BO877217 ACBNCOURT
38	19	5.5	1271	10	BB215199	BB215199 BB215199
39	18	5.2	279	10	BB380837	BB380837 BB380837
40	18	5.2	283	10	BB324963	BB324963 BB324963
41	18	5.2	294	10	BB043670	BB043670 BB043670
42	18	5.2	298	12	BG056101	BG056101 nat87611
43	18	5.2	331	9	AA659731	AA659731 nv01e10.s
44	18	5.2	331	10	BB646253	BB646253 7e83902.x
45	18	5.2	344	9	AU054035	AU054035 AU054035

ALIGNMENTS

RESULT 1
AT605049/c 151 bp mRNA linear EST 21-APR-1999
LOCUS
DEFINITION
3' mRNA sequence.

ACCESSION
AT605049
VERSION
AT605049.1 GI:4614216

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

Mus musculus
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
EST.
1 (bases 1 to 151)
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schur, R.,
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R., and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:498955

Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system -384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES

source

Location/Qualifiers

1..457
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="7120495D03"
/clone_lib="RIKEN full-length enriched, 2 days neonate
sympathetic ganglion"
/sex="mixed"
/tissue_type="sympathetic ganglion"
/dev_stage="2 days neonate"
/lab_host="DH10B"
/note="Site_1: Salt; Site_2: BamHI. cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGACGCTCTTTTCTTTTCTTTTCTTTT 3'] cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATCTCGAGTTAATTAATTAATTCACCCCTCCCTCC 3']. cDNA
was cleaved with XhoI and BamHI. Vector: a modified
pBluescript KS(+) after bulk excision from lambda FLG I.
Cloning sites, 5' end: Salt; 3' end: BamHI. Host: DH10B.
-RNA was provided by Akira Nakagawa, Div. of
Biochemistry, Chiba Cancer Center Research Institute,
666-2 Nitona, Chuo-ku, Chiba, 260-8717 Japan, whose
assistance we gratefully acknowledge."

BASE COUNT

135 a 87 c 79 g 156 t

ORIGIN

Query Match 5.8%; Score 20; DB 10; Length 457;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 AAGAAAGAACCAAGAAC 42

|||||

Db 21 AAGAAAGAACCAAGAAC 2

RESULT 4

AM563834

LOCUS 576 bp mRNA linear EST 19-JUL-2000
DEFINITION LG1_272_A11.B1.A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
sequence.

ACCESSION

AM563834

VERSION

AM563834.1

KEYWORDS

EST

SOURCE

ORGANISM

Sorghum

REFERENCE

AUTHORS

TITLE

JOURNAL

Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
clade; Panicoidae; Andropogoneae; Sorghum.
1 (bases 1 to 576)
Cordonier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L.H.
An EST database from Sorghum: light-grown seedlings
Unpublished (2000)

COMMENT

Contact: Cordonier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mpm@ga.uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 574
POLYA=No.

FEATURES

source

Location/Qualifiers

1..576
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)
seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
: The library was made from poly-A RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT

84 a 202 c 174 g 116 t

ORIGIN

Query Match 5.8%; Score 20; DB 10; Length 576;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 CCCGCGGCTCTGCGCGT 315

|||||

Db 412 CCCGCGGCTCTGCGCGT 431

RESULT 5

BG310307/c

LOCUS

DEFINITION BG310307 909 bp mRNA linear EST 22-OCT-2001
HVSMEC0017113f Hordeum vulgare seedling shoot EST library
HVSMEC0017113f (Etiolated and unstressed) Hordeum vulgare cDNA clone

ACCESSION

BG310307

VERSION

BG310307.2

KEYWORDS

EST

SOURCE

ORGANISM

Hordeum vulgare

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rw@ingclemson.edu
Total hg bases = 258
Seq primer: AATTAACCTCCTCAAGG
High quality sequence start: 8
High quality sequence stop: 909.
Location/Qualifiers
1..909
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"

FEATURES

source

```

/clone="HVSMEC0017113f"
/clone_lib="Hordeum vulgare seedling shoot EST library
HYCDNA0003 (Etisolated and unstressed)"
/tissue_type="Seedling shoot"
/lab_host="TJC121"
/Note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under aseptic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedling shoots were
then harvested, total RNA was prepared, poly(A) RNA was
purified, one primary unamplified cDNA library was made,
and 1 million pfu were in vivo excised to give plusescript
SK(-) cDNA phagemids. These steps were performed in the TJ
Riverside (Choi, Close, Fenton). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wang). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wang, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wang R, Kleinholz A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html)"
BASE COUNT      266 a      181 c      343 g      112 t
ORIGIN

Query Match      5.8%; Score 20; DB 12; Length 909;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GAAACACCAACCCCGCCAC 57
|||||
Db 388 GAAACACCAACCCCGCCAC 369

RESULT 6
BE179413
LOCUS      925 bp      mRNA      linear      EST 31-OCT-2000
601806981F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4037661 5',
mRNA sequence.
ACCESSION  BE179413
VERSION     BE179413.1 GI:11057555
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 925)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: L1AM9315 row: a column: 22
            High quality sequence stop: 659.
            Location/Qualifiers
            1..925
            /organism="Mus musculus"
            /strain="C57BL/6J"

FEATURES
source

/clone="HVSMEC0017113f"
/clone_lib="Hordeum vulgare seedling shoot EST library
HYCDNA0003 (Etisolated and unstressed)"
/tissue_type="Seedling shoot"
/lab_host="TJC121"
/Note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under aseptic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedling shoots were
then harvested, total RNA was prepared, poly(A) RNA was
purified, one primary unamplified cDNA library was made,
and 1 million pfu were in vivo excised to give plusescript
SK(-) cDNA phagemids. These steps were performed in the TJ
Riverside (Choi, Close, Fenton). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wang). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wang, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wang R, Kleinholz A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html)"
BASE COUNT      266 a      181 c      343 g      112 t
ORIGIN

Query Match      5.8%; Score 20; DB 12; Length 909;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GAAACACCAACCCCGCCAC 57
|||||
Db 388 GAAACACCAACCCCGCCAC 369

RESULT 7
AA853727
LOCUS      222 bp      mRNA      linear      EST 20-JUN-2002
NHTBCae07h07r1 Normal Human Trabecular Bone Cells Homo sapiens CDNA
clone NHTBCae07h07, mRNA sequence.
ACCESSION  AA853727
VERSION     AA853727.1 GI:2940466
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 222)
AUTHORS   Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G.,
            Beckstrom-Sternberg,S.M., Green,E.D., Powell,J.I., Yang,L.M., Robey
            ,P.G., Hotchkiss,R.N. and Francomano,C.A.
            SGAP: The Skeletal Genome Anatomy Project
            Unpublished (1997)
            Contact: Libbin Jia
            Medical Genetics Branch
            National Human Genome Research Institute
            10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
            Tel: 301-402-4877
            Fax: 301-496-7157
            Email: libbin@helix.nih.gov
            Seq primer: M13 Reverse
            Location/Qualifiers
            1..222
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="NHTBCae07h07"
            /clone_lib="Normal Human Trabecular Bone Cells"
            /sex="Female"
            /tissue_type="Bone"
            /cell_type="Trabecular Bone Cells"
            /lab_host="SURE"
            /note="Organ: Hip; Vector: pBluescript; Site 1: EcoRI;
            library constructed by Dr. Marian Young and Dr. Pamela
            Gehron Robey (NIHCR)"
BASE COUNT      33 a      49 c      76 g      64 t
ORIGIN

Query Match      5.5%; Score 19; DB 9; Length 222;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 GGCAGGTCTCTGGGCTCAGC 234
|||||
Db 92 GGCAGGTCTCTGGGCTCAGC 110

```

RESULT 8
 AZ239458/c
 LOCUS
 DEFINITION RPCI-23-73L2.TV RPCI-23 Mus musculus genomic clone RPCI-23-73L2.
 DNA sequence.
 ACCESSION AZ239458
 VERSION A2239458.1 GI:8552649
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 240)
 Z hao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akninet, B., Levin, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-23
 JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: RPCI-23-73L2.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)
 plate: 73 row: L column: 2
 Seq primer: T7
 Class: BAC ends.
 FEATURES
 Location/Qualifiers
 source
 1..240
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-73L2"
 /clone_1id="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain: Vector: pBACe3.6; Site: 1: EcoRI, Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 66 a 30 c 27 g 117 t
 ORIGIN
 Query Match 5.5%; Score 19; DB 17; Length 240;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 23 AAGAAACCAAGAGAA 41
 ||||||||||||||||
 Db 238 AAGAAACCAAGAGAA 220
 RESULT 9
 D79948
 LOCUS
 DEFINITION HUM362E05B Human aorta polyA+ (Tfujiwara) Homo sapiens cDNA clone
 GEN:362E05 5', mRNA sequence.
 D79948
 VERSION D79948.1 GI:1180299
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 38)
 Fujiiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takachi, A., Takada, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S., and Nakamura, Y.
 TITLE Fujiiwara et al. (1995)
 JOURNAL Unpublished (1995)
 COMMENT Contact: Tsutomu Fujiiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno, Kawasuchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.
 FEATURES
 Location/Qualifiers
 source
 1..38
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GEN-362E05"
 /clone_1id="Human aorta polyA+ (Tfujiwara)"
 /tissue_type="aorta"
 /note="Organ: brain"
 BASE COUNT 119 a 58 c 51 g 107 t 3 others
 ORIGIN
 Query Match 5.5%; Score 19; DB 14; Length 338;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 25 AGAAACCAAGAGAA 43
 ||||||||||||||||
 Db 183 AGAAACCAAGAGAA 201
 RESULT 10
 B1307759/c
 LOCUS
 DEFINITION TY1D03 hepatocellular carcinoma expression library Marmota monax
 CDNA clone TY1D03, mRNA sequence.
 B1307759
 VERSION B1307759.1 GI:14982084
 KEYWORDS EST.
 SOURCE woodchuck.
 ORGANISM Marmota monax
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae; Marmota.
 1 (bases 1 to 402)
 Yamamoto, T., Hazel, C., Yerges, L., and Mason, W.S.
 Hepatocellular carcinoma/liver differential expression libraries
 Unpublished (2001)
 COMMENT Contact: Mason WS
 Fox Chase Cancer Center
 7701 Burholme Avenue, Philadelphia, PA 19111, USA
 Tel: 215 728 2462
 Fax: 215 728 3105
 Email: ws_mason@fccc.edu
 Insert Length: 402 Std Error: 0.00
 High quality sequence stop: 402.
 FEATURES
 Location/Qualifiers
 source
 1..402
 /organism="Marmota monax"
 /db_xref="taxon:9995"
 /clone="TY1D03"
 /clone_1id="hepatocellular carcinoma expression library"
 /tissue_type="Hepatocellular carcinoma"
 /note="Organ: Liver; cDNAs were synthesized by reverse transcription of mRNAs purified from the liver and a hepatocellular carcinoma (HCC) of a woodchuck (marmota monax) chronically infected with woodchuck hepatitis virus. cDNAs were digested with RsaI and ligated to one

of two different adaptors. HCC enriched cDNA was then generated using the Clontech PCR Select cDNA subtraction kit. Amplicons with non-identical adaptors at each end were selectively amplified by PCR, using primers specific to the adaptors, and the resulting amplicons were ligated into pGEM T easy (Promega)."

BASE COUNT 107 a 100 c 96 g 99 t
ORIGIN

Query Match 5.5%; Score 19; DB 13; Length 402;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 261 AATGAGGCGCGGCTGGG 279
Db 179 AATGAGGCGTGGG 161

RESULT 11 457 bp mRNA linear EST 10-OCT-2001
LOCUS BB679924 RIKEN full-length enriched, adult male testis Mus musculus
DEFINITION BB679924 cDNA clone 4922505P09 3', mRNA sequence.
ACCESSION BB679924
VERSION BB679924.1 GI:1600657
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 457)
AUTHORS Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akibira,S., Tanaka,T., Tomaru,A., Toya,T., Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished (2001)
COMMENT Contact: Yoshitake Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp.
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
waij,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuyama,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES
Location/Qualifiers
1..457
Source

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4922505P09"
/clone_1lb="RIKEN full-length enriched, adult male testis"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGACCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGCGCCATTAATTCGAGTTAATTAATTAATTCCTCCCTCC 3']. cDNA was cloned into the XhoI and BamHI sites."

BASE COUNT 175 a 105 c 63 g 114 t
ORIGIN

Oy 23 AAGGAAACCAAGGAGG 41
Db 224 AAGGAAACCAAGGAGG 242

RESULT 12 469 bp mRNA linear EST 22-JUN-2000
LOCUS AM200309/c
DEFINITION da18e02.y1 normalized Xenopus laevis gastrula Xenopus laevis cDNA
ACCESSION AM200309
VERSION AM200309.1 GI:6480878
KEYWORDS EST.
SOURCE African Clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 469)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterson,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Bruce Song
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for this library can be found through Research Genetics, visit their web page at: http://www.resgen.com/
Seq primer: -40RP from Gldco
High quality sequence stop: 429.
Location/Qualifiers
1..469
/organism="Xenopus laevis"

FEATURES
Location/Qualifiers
1..469
Source

```

/db_xref="taxon:8355"
/clone="XENOPUS_SOURCE.ID:xinga002g03"
/clone_lib="normalized xenopus laevis gastrula"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from 2ug of poly A+ RNA (equal
parts from stage 10.5 and stage 11.5 gastrulae).
EcoRI-XhoI cut cDNA was then ligated into Unizap-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the 3'
end. SS-library phagemids were prepared by mass excision
from the original library and normalized by hybridization
to biotinylated driver (prepared from the same library by
PCR) to Cot-omega of 11. After removal of hybrids and
excess driver by streptavidin sepharose chromatography,
the ss-phagemids were made double stranded and
electroporated into Top-10 F'. Original library
construction by Bruce Blumberg (Cho et al. 1991 Cell 67,
111-1120). Normalized by Jihwan Song (Song, Cho and
Blumberg, unpublished). Note: This is a Xenopus Gene
Collection (XGC) library."
BASE COUNT      136 a      86 c      60 g      187 t
ORIGIN
Query Match      5.5%; Score 19; DB 10; Length 469;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 15 TAACGACCAAGAAAACC 33
|||||
Db 301 TAACGACCAAGAAAACC 283

RESULT 13
BO789910/c 471 bp mRNA linear EST 30-JUL-2002
LOCUS
DEFINITION
hage004ab10 Heterobasidion annosum - Scots pine infection stage
(HAGE) subtraction cDNA library Pinus sylvestris/Heterobasidion
annosum cDNA clone hage004ab10, mRNA sequence.
ACCESSION
BO789910.1 GI:22004872
VERSION
EST.
SOURCE
Pinus sylvestris/Heterobasidion annosum.
ORGANISM
Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE
1 (bases 1 to 471)
Asiegbu, F.O., Nahalkova, J. and Dean, R.A.
Selected expressed sequence tags of cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
unpublished (2001)
JOURNAL
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: 17 primer.
FEATURES
source
Location/Qualifiers
1..471
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hage004ab10"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage (HAGE) subtraction cDNA library"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/notes="Vector: pT-Adv; Site_1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots pine
roots infected for 6-days with mycelia of Heterobasidion
annosum (FP5)."
```

```

ORIGIN
Query Match      5.5%; Score 19; DB 14; Length 471;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 23 AAAGAAAACCAAGAAA 41
|||||
Db 74 AAAGAAAACCAAGAAA 56

RESULT 14
BG569327/c 524 bp mRNA linear EST 10-APR-2001
LOCUS
DEFINITION
60258863.F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722752 5',
mRNA sequence.
ACCESSION
BG569327.1 GI:13576980
VERSION
BG569327
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 524)
NIH-MGC http://mgc.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapds-retail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: ILICM1581 row: c column: 09
High quality sequence stop: 524.
location/Qualifiers
1..524
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4722752"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctggcc); Site_2: SfiI (ggccattggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCAATATGAGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGCGCGCATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
```

```

BASE COUNT      126 a      101 c      119 g      125 t
FEATURES
source
Location/Qualifiers
1..471
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hage004ab10"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage (HAGE) subtraction cDNA library"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/notes="Vector: pT-Adv; Site_1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots pine
roots infected for 6-days with mycelia of Heterobasidion
annosum (FP5)."
```

```

BASE COUNT      149 a      76 c      101 g      198 t
ORIGIN
Query Match      5.5%; Score 19; DB 12; Length 524;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 23 AAAGAAAACCAAGAAA 41
|||||
Db 394 AAAGAAAACCAAGAAA 376

RESULT 15
AO799264/c 540 bp DNA linear GSS 09-AUG-1999
LOCUS
DEFINITION
HS_5451_B1.F05.J7A RPO1-11 Human MALE BAC library Homo sapiens
genomic clone Plate=1027 Col=9 Row=L, DNA sequence.
ACCESSION
AO799264
```

```

VERSION      AQ799264.1  GI:5716596
KEYWORDS
SOURCE
ORGANISM     human.
              Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 540)
AUTHORS      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
              Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
              Hood,L.
TITLE        Sequence-tagged connectors: A sequence approach to mapping and
              scanning the human genome
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE      99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
              High Throughput Sequencing Center
              University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
              Email: jwallace@u.washington.edu
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieter@dejong.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)
              or from Research Genetics (info@resgen.com). BAC end Web Server:
              http://www.htsc.washington.edu
              Plate: 1027 row: L column: 9
              Seq primer: 77
              Class: BAC ends
              High quality sequence stop: 540.
FEATURES
  source
    1..540
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="Plate=1027 Col=9 Row=L"
        /clone_lib="RPCI-11 Human Male BAC Library"
        /sex="male"
        /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
        Male blood DNA was isolated from one randomly chosen donor
        and partially digested with a combination of EcoRI and
        .EcoRI Methylase. Size selected DNA was cloned into the
        pBACe3.6 vector at EcoRI sites"
BASE COUNT   169 a      89 c      93 g      173 t      16 others
ORIGIN
Query Match      5.5%; Score 19; DB 17; Length 540;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      26 GAAAAACCAAGAAACAC 44
        ||||||||||||||||
Db      322 GAAAAACCAAGAAACAC 304

```

Search completed: February 19, 2003, 02:07:41
 Job time : 1467 secs

Thu Feb 20 14:41:26 2003

us-09-873-224-147.oli.rml

Page 1

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 01:16:20 ; Search time 75 seconds
(without alignments)
1410.713 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 345
Sequence: 1 ATGAGCAGACTTCTTAACC.....AATGACCCCGGCGCAGGA 345

Scoring table: OLIGO_NDC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size: 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
4: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCUS.COMB.seq:*
6: /cgn2_6/prodata/1/ina/ackfileseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	309	89.6	309	4	US-08-836-075A-49
2	43	12.5	549	3	US-08-441-971-60
3	43	12.5	549	4	US-08-221-653-60
4	43	12.5	549	4	US-08-442-144A-60
5	43	12.5	549	4	US-08-441-970-60
6	43	12.5	573	2	US-08-290-665A-141
7	43	12.5	573	2	PCT-US95-10398-141
8	43	12.5	573	4	US-08-836-075A-65
9	40	11.6	573	2	US-08-290-665A-142
10	40	11.6	573	2	PCT-US95-10398-142
11	38	11.0	573	2	US-08-290-665A-136
12	38	11.0	573	2	PCT-US95-10398-136
13	35	10.1	573	2	US-08-290-665A-137
14	35	10.1	573	2	US-08-290-665A-138
15	35	10.1	573	2	PCT-US95-10398-137
16	35	10.1	573	2	PCT-US95-10398-138
17	35	10.1	573	5	PCT-US95-10398-139
18	35	10.1	573	5	PCT-US95-10398-139
19	35	10.1	803	1	US-08-157-235-1
20	35	10.1	803	1	US-08-157-235-2
21	35	10.1	803	1	US-08-157-235-3
22	35	10.1	803	1	US-08-157-235-4
23	34	9.9	573	2	US-08-290-665A-135
24	34	9.9	573	2	PCT-US95-10398-135
25	34	9.9	803	1	US-08-157-235-5
26	31	9.0	183	1	US-07-681-703B-21
27	31	9.0	183	2	US-08-407-410B-21

28	31	9.0	183	2	US-08-485-500-21	Sequence 21, Appl
29	31	9.0	183	5	PCT-US91-02370-21	Sequence 21, Appl
30	31	9.0	270	1	US-07-681-703B-23	Sequence 23, Appl
31	31	9.0	270	2	US-08-407-410B-23	Sequence 23, Appl
32	31	9.0	270	2	US-08-485-500-23	Sequence 23, Appl
33	31	9.0	270	5	PCT-US91-02370-23	Sequence 23, Appl
34	31	9.0	273	1	US-07-681-703B-19	Sequence 19, Appl
35	31	9.0	273	2	US-08-407-410B-19	Sequence 19, Appl
36	31	9.0	273	2	US-08-485-500-19	Sequence 19, Appl
37	31	9.0	273	2	PCT-US91-02370-19	Sequence 19, Appl
38	31	9.0	306	2	US-08-537-811-35	Sequence 35, Appl
39	31	9.0	327	4	US-08-836-075A-1	Sequence 1, Appl
40	31	9.0	355	3	US-08-444-818-104	Sequence 104, App
41	31	9.0	355	3	US-08-444-818-106	Sequence 106, App
42	31	9.0	360	1	US-07-681-703B-17	Sequence 17, Appl
43	31	9.0	360	2	US-08-407-410B-17	Sequence 17, Appl
44	31	9.0	360	2	US-08-485-500-17	Sequence 17, Appl
45	31	9.0	360	5	PCT-US91-02370-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-836-075A-49
Sequence 49, Application US/08836075A
Patent No. 6180768
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
APPLICANT: STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,075A
FILING DATE: 21 Apr 1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS: 004
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 309 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-836-075A-49

RESULT 2
US-08-441-971-60

Patent No. 6071693
GENERAL INFORMATION:

APPLICANT: Tai-An Cha
TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston

STATE: Massachusetts
COUNTRY: USA
ZIP: 02210

COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS Version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
ADDRESS:

APPLICATION NUMBER: US/08/441,971
FILING DATE: 16-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/221,653
FILING DATE:
APPLICATION NUMBER: US/07/881,528
FILING DATE:

APPLICATION NUMBER: 07/697,326
FILING DATE: 8 May 1991
ATTORNEY/AGENT INFORMATION:
NAME: Janiuk, Anthony J.
REGISTRATION NUMBER: 30,000

NOTIFICATION NUMBER: 29,809
REFERENCE/DOCKET NUMBER: C0772/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 720-3558

TELEPHONE: (617) 720-3500
TELEFAX: (617) 720-2441
TELEX: EZEKIEL
INFORMATION FOR SEC ID NO. 50

SEQUENCE CHARACTERISTICS:
LENGTH: 549 nucleotides
TYPE: nucleic acid

```

; STRINDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: nac5
US-08-441-971-60

```

Query Match	12.5%;	Score 43;	DB 3;	Length 549;
Best Local Similarity	100.0%;	Pred. No. 2e-12;		
Matches 43;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	211	CCGAGGGCAGTCTCTGGGCTACGCCCGGGTACCCTTGCCCCCT	25
Db	212	CCGAGGGCAGTCTCTGGGCTACGCCCGGGTACCCTTGCCCCCT	25

RESULT 3
US-08-221-653-60
: Sequence 60

Sequence 80, Application US/08221653
Patent No. 6190864

GENERAL INFORMATION:
APPLICANT: Tai-An Cha
TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESS:

ADDRESSEE: WOLT, Greenfield & Sacks, P.C
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02210

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS Version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/221,652
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/881,528
FILING DATE:

FILING DATE:
 APPLICATION NUMBER: 07/697,326
 FILING DATE: 8 May 1991
 ATTORNEY/AGENT INFORMATION:

NAME: Janiuk, Anthony J.
REGISTRATION NUMBER: 29, 809
REFERENCE/DOCKET NUMBER: C0772/7000
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 720-3500
TELEFAX: (617) 720-2441
TELEX: EZEKIEL

INFORMATION FOR SEQ ID NO:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 549 nucleotides
 TYPE: nucleic acid

STRANDEDNESS: si
TOPOLOGY: linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE:

ORIGINAL SOURCE: nacc5
INDIVIDUAL ISOLATE: 5-08-221-653-60

```
Query Match      12.5%; Score 43; DB 4; Length 549;
Best Local Similarity 100.0%; Pred. No. 2e-12;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

211 CCGAGGCGAGGTCTCTGCGGCTCAGCCCGGGTACCTTGGCCCT 25
|||||
212 CCGAGGCGAGGTCTCTGCGGCTCAGCCCGGGTACCTTGGCCCT 25

RESULT 4

US-08-442-144A-60

Sequence 60, Application US/08442144A
Patent No. 6214583

GENERAL INFORMATION:

APPLICANT: Tai-An Cha
APPLICANT: Eileen Beall
APPLICANT: Bruce Irvine
APPLICANT: Janice Kolberg
APPLICANT: Michael S. Urdea
TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 Inch
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows NT
SOFTWARE: Microsoft Word 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/442,144A
FILING DATE: MAY 16, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/221,653

FILING DATE: APRIL 1, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Doreen Yalko Trujillo

REGISTRATION NUMBER: 35,719

REFERENCE/DOCKET NUMBER: CHIR-0121

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

TELEX:

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 549 Nucleotides

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

MOLECULE TYPE: DNA

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: nac5

US-08-442-144A-60

Query Match 12.5%; Score 43; DB 4; Length 549;
Best Local Similarity 100.0%; Pred. No. 2e-12;

Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CCGAGGCGAGGTCCTGGGCTCAGCCCGGTACCCCTTGCCCT 253

Db 212 CCGAGGCGAGGTCCTGGGCTCAGCCCGGTACCCCTTGCCCT 254

RESULT 5

US-08-441-970-60

Sequence 60, Application US/08441970

Patent No. 6297370

GENERAL INFORMATION:

APPLICANT: Tai-An Cha
TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 Inch

COMPUTER: IBM compatible

OPERATING SYSTEM: MS-DOS Version 3.3

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/441,970

FILING DATE: 16-MAY-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/881,528

FILING DATE: 08-MAY-1992

APPLICATION NUMBER: 07/697,326

FILING DATE: 8 May 1991

ATTORNEY/AGENT INFORMATION:

NAME: Janluk, Anthony J.

REGISTRATION NUMBER: 29,809

REFERENCE/DOCKET NUMBER: C0772/7000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 720-3500

TELEFAX: (617) 720-2441

TELEX: EZEKTEL

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 549 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: nac5

US-08-441-970-60

QY 211 CCGAGGCGAGGTCCTGGGCTCAGCCCGGTACCCCTTGCCCT 253

Db 212 CCGAGGCGAGGTCCTGGGCTCAGCCCGGTACCCCTTGCCCT 254

RESULT 6

US-08-290-665A-141

Sequence 141, Application US/08290665A

Patent No. 3682852

GENERAL INFORMATION:

APPLICANT: BURK, J., MILLER, R.H. AND
APPLICANT: BURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DECODED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A

FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ. ID NO.: 141:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z1
US-08-290-665A-141

Query Match 12.5%; Score 43; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. No. 2e-12;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CCGAGGGCAGTCTGGGCTCAGCCCGGTAACCTTGCCCCCT 253
DB 212 CCGAGGGCAGTCTGGGCTCAGCCCGGTAACCTTGCCCCCT 254

RESULT 7

PCT-US95-10398-141
Sequence 141, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BORK, J., MILLER, R.H. AND
PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO.: 141:

SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z1
PCT-US95-10398-141

Query Match 12.5%; Score 43; DB 5; Length 573;
Best Local Similarity 100.0%; Pred. No. 2e-12;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CCGAGGGCAGTCTGGGCTCAGCCCGGTAACCTTGCCCCCT 253
DB 212 CCGAGGGCAGTCTGGGCTCAGCCCGGTAACCTTGCCCCCT 254

RESULT 8

US-08-836-075A-65
Sequence 65, Application US/08836075A
Patent No. 6180768
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEBERT
STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
TITLE OF INVENTION: AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,075A
FILING DATE: 21 Apr 1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO.: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 831 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-836-075A-65

Query Match 12.5%; Score 43; DB 4; Length 831;
Best Local Similarity 100.0%; Pred. No. 2e-12;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CCGAGGCGAGTCTGGGCTACGCCGGTACCTTTGGCCCT 253
DB 227 CCGAGGCGAGTCTGGGCTACGCCGGTACCTTTGGCCCT 269

RESULT 9

US-08-290-665A-142
Sequence 142, Application US/08290665A
Patent No. 5882852

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R.H. AND
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 142:

SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 25
US-08-290-665A-142

Query Match 11.6%; Score 40; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. No. 6.2e-11;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CCGAGGCGAGTCTGGGCTACGCCGGTACCTTTGGCC 250
DB 212 CCGAGGCGAGTCTGGGCTACGCCGGTACCTTTGGCC 251

RESULT 10

PCT-US95-10398-142

Sequence 142, Application PC/TUS9510398
GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R.H. AND
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 142:

SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 25
PCT-US95-10398-142

Query Match 11.6%; Score 40; DB 5; Length 573;
Best Local Similarity 100.0%; Pred. No. 6.2e-11;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CCGAGGCGAGTCTGGGCTACGCCGGTACCTTTGGCC 250
DB 212 CCGAGGCGAGTCTGGGCTACGCCGGTACCTTTGGCC 251

RESULT 11

US-08-290-665A-136
Sequence 136, Application US/08290665A
Patent No. 5882852

GENERAL INFORMATION:

APPLICANT: BURKH, J., MILLER, R.H. AND
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 136:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S52
US-08-290-665A-136

Query Match 11.0%; Score 38; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. No. 6.2e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 AATGAGGCTGGCGGTGGCGAGGTGCTCTGTCCTCC 298
DB 262 AATGAGGCTGGCGGTGGCGAGGTGCTCTGTCCTCC 299

RESULT 12
PCT-US95-10398-136
Sequence 136, Application PC/TUS9510398
GENERAL INFORMATION:
APPLICANT: BORK, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10398
FILING DATE: 15-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/086,428
FILING DATE: 29 JUNE 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290/665
FILING DATE: 15 AUGUST 1994
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 136:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S52
PCT-US95-10398-136

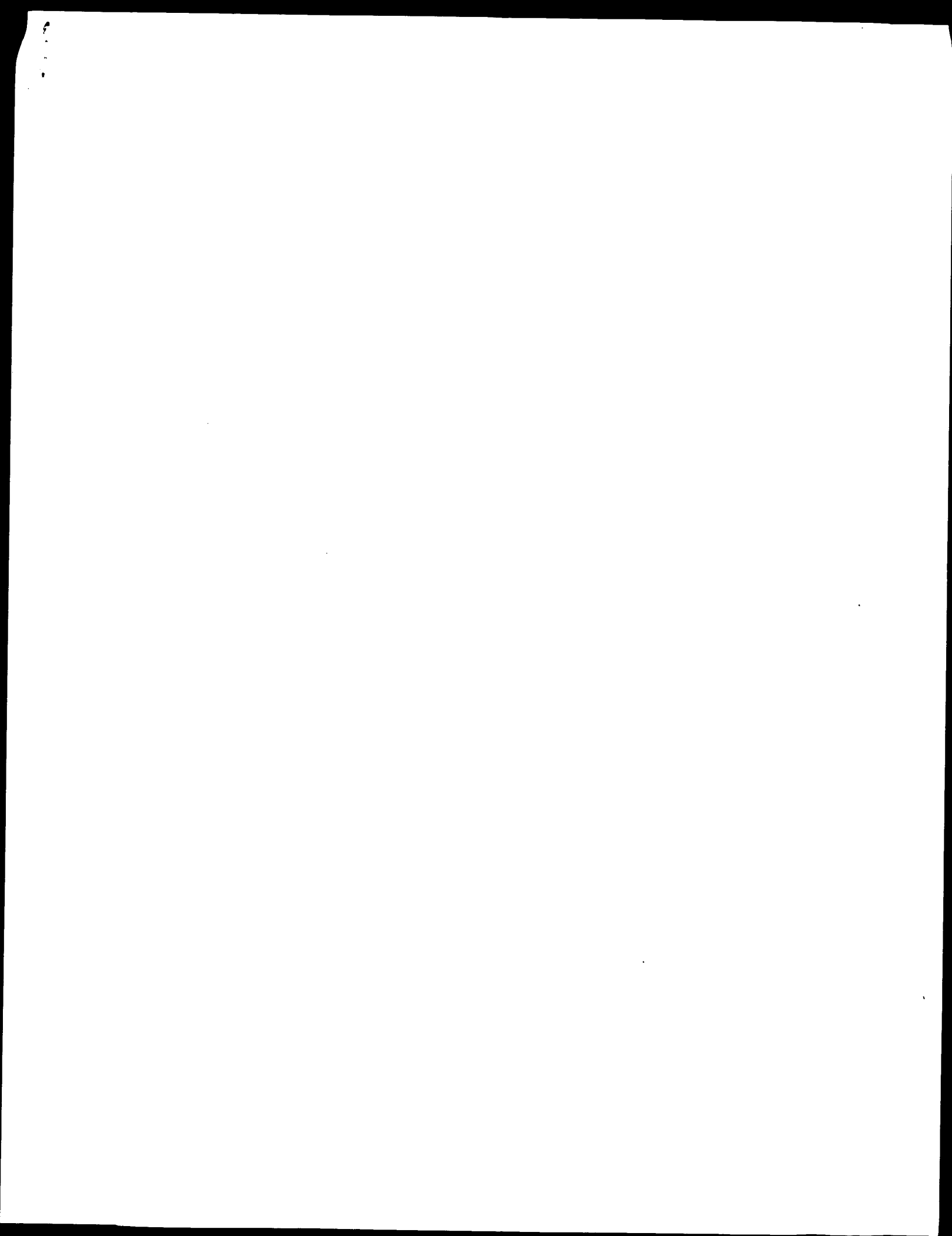
Query Match 11.0%; Score 38; DB 5; Length 573;
Best Local Similarity 100.0%; Pred. No. 6.2e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 AATGAGGCTGGCGGTGGCGAGGTGCTCTGTCCTCC 298
DB 262 AATGAGGCTGGCGGTGGCGAGGTGCTCTGTCCTCC 299

RESULT 13
US-08-290-665A-137
Sequence 137, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BORK, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 137:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
US-08-290-665A-137

Query Match 10.1%; Score 35; DB 2; Length 573;

Search completed: February 19, 2003, 02:09:07
Job time : 79 secs



Thu Feb 20 14:41:27 2003

us-09-873-224-147.oli.rnpb

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 19, 2003, 01:18:25 ; Search time 105 Seconds
(without alignments)
1673.492 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 345
Sequence: 1 ATGAGCACACTCTTAAC.....AAATGACCCCGCGCAGGA 345

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 424239 seqs, 25461826 residues

Word size: 0
Total number of hits satisfying chosen parameters: 848478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Published Applications-NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCR_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	309	89.6	309	US-09-851-138-49	Sequence 49, Appl
2	296	85.8	346	US-09-899-046-147	Sequence 147, Appl
3	43	12.5	831	US-09-851-138-65	Sequence 65, Appl
4	31	9.0	152	US-09-921-387-39	Sequence 39, Appl
5	31	9.0	234	US-09-921-387-41	Sequence 41, Appl
6	31	9.0	310	US-09-921-387-114	Sequence 114, Appl
7	31	9.0	327	US-09-851-138-1	Sequence 1, Appl
8	31	9.0	339	US-09-921-397-115	Sequence 115, Appl
9	31	9.0	2433	US-09-973-025-49	Sequence 49, Appl
10	31	9.0	9379	US-09-916-359-1	Sequence 1, Appl
11	31	9.0	9416	US-10-104-966-13	Sequence 13, Appl
12	31	9.0	9416	US-09-995-937-19	Sequence 19, Appl
13	31	9.0	9416	US-09-928-076-19	Sequence 19, Appl
14	31	9.0	9416	US-09-928-955-13	Sequence 13, Appl
15	31	9.0	9646	US-09-995-937-1	Sequence 1, Appl
16	31	9.0	9646	US-09-742-659-3	Sequence 3, Appl
17	31	9.0	9646	US-09-238-076-1	Sequence 1, Appl
18	31	9.0	10803	US-09-747-419-17	Sequence 17, Appl
19	31	9.0	12980	US-09-995-937-5	Sequence 5, Appl

20	31	9.0	12980	US-09-238-076-5	Sequence 5, Appl
21	28	8.1	223	US-09-851-138-9	Sequence 9, Appl
22	28	8.1	499	US-09-899-046-151	Sequence 151, Appl
23	28	8.1	509	US-09-899-046-41	Sequence 41, Appl
24	28	8.1	509	US-09-899-046-43	Sequence 43, Appl
25	28	8.1	652	US-09-851-138-59	Sequence 59, Appl
26	28	8.1	959	US-09-899-046-49	Sequence 49, Appl
27	28	8.1	959	US-09-899-046-51	Sequence 51, Appl
28	28	8.1	959	US-09-899-046-53	Sequence 53, Appl
29	28	7.8	153	US-09-736-959A-30	Sequence 30, Appl
30	27	7.8	413	US-09-851-138-45	Sequence 45, Appl
31	27	7.8	630	US-09-973-322-1	Sequence 1, Appl
32	27	7.8	957	US-09-851-138-11	Sequence 11, Appl
33	27	7.8	1195	US-09-851-138-75	Sequence 75, Appl
34	27	7.8	2499	US-09-881-239-4	Sequence 4, Appl
35	26	7.5	288	US-09-899-046-191	Sequence 191, Appl
36	26	7.5	499	US-09-899-046-165	Sequence 165, Appl
37	26	7.5	499	US-09-899-046-163	Sequence 163, Appl
38	25	7.2	957	US-09-851-138-43	Sequence 43, Appl
39	25	7.2	957	US-09-851-138-5	Sequence 5, Appl
40	24	7.0	310	US-09-851-138-21	Sequence 21, Appl
41	23	6.7	498	US-09-899-046-193	Sequence 193, Appl
42	23	6.7	957	US-09-851-138-17	Sequence 17, Appl
43	23	6.7	957	US-09-899-046-143	Sequence 143, Appl
44	23	6.1	51	US-09-756-875-4	Sequence 4, Appl
45	21	6.1	10	US-09-756-875-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-851-138-49
Sequence 49, Application US/09851138
Publication No. US20020183508A1
GENERAL INFORMATION:
APPLICANT: MAERRENS, GEERT
STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESS: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION/DOCKET NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 309 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 49;
US-09-851-138-49

Query Match
Best Local Similarity 89.6%; Score 309; DB 9; Length 309;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAGGACACTTCTTAACCAACCAAGAAAAACCAAAACCAACACCAACCCGCGACAG 60
DB 1 ATAGGACACTTCTTAACCAACCAAGAAAAACCAAAACCAACACCAACCCGCGACAG 60
QY 61 AGTTAACTTCCAGAGCGCGCGTGCAGATCGTTGGAGTTTACGTCTACCAACCGAGG 60
DB 61 AGTTAACTTCCAGAGCGCGCGTGCAGATCGTTGGAGTTTACGTCTACCAACCGAGG 60
QY 121 GCCCCAGTTGGGTGTCGTGCAGTGGCAAGACTTCCAGAGCGGTCCAGACCTGCGAGTA 120
DB 121 GCCCCAGTTGGGTGTCGTGCAGTGGCAAGACTTCCAGAGCGGTCCAGACCTGCGAGTA 120
QY 181 GGGCCCAACCCATCCCAAGGCGCGCGCAACCGAGGAGGTCTGCGGCTGACGCGGCT 180
DB 181 GGGCCCAACCCATCCCAAGGCGCGCGCAACCGAGGAGGTCTGCGGCTGACGCGGCT 180
QY 241 ACCCTTGGCCCTATATGGAATGAGGCGTGGGCTGGGCGAGGCTGCTCTGCTCCCGG 240
DB 241 ACCCTTGGCCCTATATGGAATGAGGCGTGGGCTGGGCGAGGCTGCTCTGCTCCCGG 240
QY 301 GCGGCTCTC 309
DB 301 GCGGCTCTC 309

RESULT 2

US-09-899-046-147
Sequence 147, Application US/09899046
Publication No. US20030008274A1
GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus
TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,046
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:

SEQUENCE CHARACTERISTICS:
LENGTH: 346 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..346
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1..342
US-09-899-046-147

Query Match
Best Local Similarity 85.8%; Score 296; DB 9; Length 346;
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 CCGCCACAGAGAGCTTAAGTTCACAGGCGCGGTGAGATCGTTGGAGTTTACGTGCT 109
DB 51 CCGCCACAGAGAGCTTAAGTTCACAGGCGCGGTGAGATCGTTGGAGTTTACGTGCT 110
QY 110 ACCACGAGGCGCGCGCACTTGGGTGTCGTGCAGTGGCAAGACTTCCAGGCGTCCGA 169
DB 111 ACCACGAGGCGCGCGCACTTGGGTGTCGTGCAGTGGCAAGACTTCCAGGCGTCCGA 170
QY 170 ACTTGCAGTAGGCGCAACCATCCAGGCGCGCGCAACCGAGGCGAGTCTGCGG 229
DB 171 ACTTGCAGTAGGCGCAACCATCCAGGCGCGCGCAACCGAGGCGAGTCTGCGG 230
QY 230 TCAGCCGCGGTACCTTGGCGCTATATGGAATGAGGCGTGGGCGAGGCGT 289
DB 231 TCAGCCGCGGTACCTTGGCGCTATATGGAATGAGGCGTGGGCGAGGCGT 290
QY 290 CCGTCCCGCGCGCGCTGCGCGCGTGGCGCGCAATGAGCCCGCGCGAGGA 345
DB 291 CCGTCCCGCGCGCGCTGCGCGCGTGGCGCGCAATGAGCCCGCGCGAGGA 346

RESULT 3

US-09-851-138-65
Sequence 65, Application US/09851138
Publication No. US20020183508A1
GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESS: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 831 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 65;
US-09-851-138-65

Query Match 12.5%; Score 43; DB 9; Length 831;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 221 CCGAGGCGAGCTCCGCGGCTACGCCCCGGGTACCTTGCCCCCT 253
DB 227 CCGAGGCGAGGCTCCGCGGCTACGCCCCGGGTACCTTGCCCCCT 269

RESULT 4

US-09-921-397-39
Sequence 39, Application US/09921397
Patent No. US20020151484A1
GENERAL INFORMATION:
APPLICANT: HYBRIGENICS
TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
TITLE OF INVENTION: applications thereof
FILE REFERENCE: B4809A - JAZ
CURRENT APPLICATION NUMBER: US/09/921,397
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: EP 00402225.7
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 156
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 39
LENGTH: 152
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-921-397-39

Query Match 9.0%; Score 31; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 4e-07;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 CCGGCTCAGCCCGGCTACCTTGCCCCCT 253
DB 120 CCGGCTCAGCCCGGCTACCTTGCCCCCT 150

RESULT 5

US-09-921-397-41
Sequence 41, Application US/09921397
Patent No. US20020151484A1
GENERAL INFORMATION:
APPLICANT: HYBRIGENICS
TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
TITLE OF INVENTION: applications thereof
FILE REFERENCE: B4809A - JAZ
CURRENT APPLICATION NUMBER: US/09/921,397
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: EP 00402225.7
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 156
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 41
LENGTH: 234
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-921-397-41

Query Match 9.0%; Score 31; DB 10; Length 234;
Best Local Similarity 100.0%; Pred. No. 4e-07;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 CCGGCTCAGCCCGGCTACCTTGCCCCCT 253
DB 186 CCGGCTCAGCCCGGCTACCTTGCCCCCT 216

RESULT 6

US-09-921-397-114
Sequence 114, Application US/09921397
Patent No. US20020151484A1
GENERAL INFORMATION:
APPLICANT: HYBRIGENICS
TITLE OF INVENTION: SID nucleic acids and polypeptides selected from a
TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
TITLE OF INVENTION: applications thereof
FILE REFERENCE: B4809A - JAZ
CURRENT APPLICATION NUMBER: US/09/921,397
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: EP 00402225.7
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 156
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 114
LENGTH: 310
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-921-397-114

Query Match 9.0%; Score 31; DB 10; Length 310;
Best Local Similarity 100.0%; Pred. No. 4.1e-07;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 CCGGCTCAGCCCGGCTACCTTGCCCCCT 253
DB 264 CCGGCTCAGCCCGGCTACCTTGCCCCCT 294

RESULT 7

US-09-851-138-1
Sequence 1, Application US/09851138
Publication No. US20020183508A1
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075
FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-851-138-1

Query Match 9.0%; Score 31; DB 9; Length 327;
Best Local Similarity 100.0%; Pred. No. 4.1e-07;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CCGAGGCGAGTCTCTGGCTACGCCGGGTA 241
|||||
DB 212 CCGAGGCGAGTCTCTGGCTACGCCGGGTA 242

RESULT 8
US-09-921-397-115

Sequence 115, Application US/09921397
Patent No. US20020151484A1
GENERAL INFORMATION:
APPLICANT: HYBRIGENICS
TITLE OF INVENTION: sid nucleic acids and polypeptides selected from a
TITLE OF INVENTION: pathogenic strain of the hepatitis C virus and
TITLE OF INVENTION: applications thereof
FILE REFERENCE: BA809A - JAZ
CURRENT APPLICATION NUMBER: US/09/921,397
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: EP 00402225.7
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 156
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 115
LENGTH: 339
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-921-397-115

Query Match 9.0%; Score 31; DB 10; Length 339;
Best Local Similarity 100.0%; Pred. No. 4.1e-07;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 CCTGGGCTACGCCGGGTACCTTGCCCT 253
|||||
DB 224 CCTGGGCTACGCCGGGTACCTTGCCCT 254

RESULT 9
US-09-973-025-49

Sequence 49, Application US/09973025
Publication No. US20020182706A1
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
BOSMAN, FONS
DE MARTYNOFF, GUY
BUYSE, MARIE-ANGE
TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE
TITLE OF INVENTION: PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE
CORRESPONDENCE ADDRESSES:
NUMBER OF SEQUENCES: 111
ADDRESS: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
APPLICATION NUMBER: US/09/973,025
FILING DATE: 10-OCT-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/612,973

FILING DATE: 11-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 1487-10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4100

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 49:

SEQUENCE CHARACTERISTICS:

LENGTH: 2433 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2430

FEATURE:

NAME/KEY: mat-peptide

LOCATION: 1..2427

SEQUENCE DESCRIPTION: SEQ ID NO: 49:

US-09-973-025-49

Query Match 9.0%; Score 31; DB 9; Length 2433;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 CCTGGGCTACGCCGGGTACCTTGCCCT 253
|||||
DB 224 CCTGGGCTACGCCGGGTACCTTGCCCT 254

RESULT 10

US-09-916-359-1

Sequence 1, Application US/09916359
Patent No. US20020034734A1
GENERAL INFORMATION:
APPLICANT: Veronique Barban
TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
TITLE OF INVENTION: TREATING C HEPATITIS
FILE REFERENCE: PMCF97-03A
CURRENT APPLICATION NUMBER: US/09/916,359
CURRENT FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: 09/388,874
PRIOR FILING DATE: 1999-09-02
PRIOR APPLICATION NUMBER: 97/02,887
PRIOR FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 9379
TYPE: DNA
ORGANISM: Virus
FEATURE:

NAME/KEY: CDS
LOCATION: (320)...(9352)
US-09-916-359-1

Query Match 9.0%; Score 31; DB 10; Length 9379;
Best Local Similarity 100.0%; Pred. No. 4.8e-07;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 CCTGGGCTACGCCGGGTACCTTGCCCT 253
|||||
DB 543 CCTGGGCTACGCCGGGTACCTTGCCCT 573

RESULT 11
US-10-104-966-13
; Sequence 13, Application US/10104966
; Patent No. US20020155124A1
; GENERAL INFORMATION:
; APPLICANT: Matti Sallberg
; APPLICANT: Catharina Hultgren
; TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
; FILE REFERENCE: TRIPEP.23AUSCI
; CURRENT APPLICATION NUMBER: US/10/104,966
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/705,547
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: 60/229,175
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 9416
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Hepatitis C virus sequence
US-10-104-966-13

Query Match
Best Local Similarity 9.0%; Score 31; DB 9; Length 9416;
Matches 31; Conservative 100.0%; Pred. No. 4.8e-07;
Indels 0; Gaps 0;

QY 223 CCTGGGCTCAGCCGGGTACCTTGCCCCCT 253
Db 565 CCTGGGCTCAGCCGGGTACCTTGCCCCCT 595

RESULT 12
US-09-995-937-19
; Sequence 19, Application US/09995937
; Publication No. US20030028010A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995,937
; FILING DATE: 28-NOV-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
US-09-929-955-13

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 19;
US-09-995-937-19

Query Match
Best Local Similarity 9.0%; Score 31; DB 9; Length 9416;
Matches 31; Conservative 100.0%; Pred. No. 4.8e-07;
Indels 0; Gaps 0;

QY 223 CCTGGGCTCAGCCGGGTACCTTGCCCCCT 253
Db 565 CCTGGGCTCAGCCGGGTACCTTGCCCCCT 595

RESULT 13
US-09-238-076-19
; Sequence 19, Application US/09238076
; Patent No. US20020102540A1
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/238,076
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/034,756
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
US-09-238-076-19

Query Match
Best Local Similarity 9.0%; Score 31; DB 10; Length 9416;
Matches 31; Conservative 100.0%; Pred. No. 4.8e-07;
Indels 0; Gaps 0;

QY 223 CCTGGGCTCAGCCGGGTACCTTGCCCCCT 253
Db 565 CCTGGGCTCAGCCGGGTACCTTGCCCCCT 595

RESULT 14
US-09-929-955-13

```

Sequence 13, Application US/09929955
Patent No. US20020136740A1
GENERAL INFORMATION:
APPLICANT: Matti Sallberg
TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
METHODS OF USE THEREOF
FILE REFERENCE: TRIPEP.23AUS2
CURRENT APPLICATION NUMBER: US/09/929,955
CURRENT FILING DATE: 2001-08-15
PRIOR APPLICATION NUMBER: 09/705,547
PRIOR FILING DATE: 2000-11-03
PRIOR APPLICATION NUMBER: 60/229,175
PRIOR FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/225,767
PRIOR FILING DATE: 2000-08-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 9416
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Hepatitis C virus sequence
US-09-929-955-13

```

```

Query Match
Best Local Similarity 100.0%; Score 31; DB 10; Length 9416;
Pred. No. 4.8e-07;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 223 CCTGGGCTCAGCCCGGTACCTTGACCCCT 253
DB 565 CCTGGGCTCAGCCCGGTACCTTGACCCCT 595

```

```

RESULT 15
US-09-995-937-1
Sequence 1, Application US/09995937
Publication No. US20030028010A1
GENERAL INFORMATION:
APPLICANT: RICE, CHARLES et al.
TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
VIRUS (HCV) AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESS: HOWELL & HAERKAMP, L.C.
STREET: 7733 FORSYTH BLVD., SUITE 1400
CITY: ST. LOUIS
STATE: MO
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/995,937
FILING DATE: 28-NO. US20030028010A1-2001
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/09/034,756
FILING DATE: 04-May-1998
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 6029-4831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9646 base pairs

```

```

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-995-937-1

```

```

Query Match
Best Local Similarity 100.0%; Score 31; DB 9; Length 9646;
Pred. No. 4.8e-07;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 223 CCTGGGCTCAGCCCGGTACCTTGACCCCT 253
DB 565 CCTGGGCTCAGCCCGGTACCTTGACCCCT 595

```

```

Search completed: February 19, 2003, 02:11:09
Job time : 112 secs

```


PF 27-APR-1994; 94WO-EP01323.
 XX
 XX 27-APR-1993; 93EP-0401099.
 PR 05-AUG-1993; 93EP-0402019.
 XX
 PA (INNO-) INNOGENETICS NV SA.
 XX
 PI Maertens G, Stuyver L;
 XX
 DR WPI: 1994-358277/44.
 DR N-PSDB: AAO78089.
 XX
 PT New polynucleotide sequences from hepatitis C virus - and related
 PT vectors, polypeptide(s) and antibodies, useful for immunisation,
 PT treatment, diagnosis and typing of HCV isolates
 XX
 PS Claim 11; Page 198-199; 404pp: English.
 XX
 CC Compositions comprising at least 5, and pref. 8 or more contiguous
 CC nucleotides selected from an HCV type 3 genomic sequence, more
 CC particularly (i) the region spanning positions 417-957 of the
 CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
 CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
 CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
 CC region spanning positions 8023-8235 of the NS5 region of the BR36
 CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
 CC sequence, may be used as primers to amplify nucleic acid from an
 CC isolate belonging to a specific genotype, or as a probe for specific
 CC detection/classification of nucleic acid. Polypeptides encoded by
 CC the nucleotides in such compositions may be used for immunisation
 CC against HCV, for the detection of antibodies directed against HCV
 CC and for serotyping. This sequence corresponds to the core region
 CC of HCV.
 XX
 SQ Sequence 115 AA:
 Alignment Scores:
 Pred. No.: 4,89e-107 Length: 115
 Score: 115.00 Matches: 115
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-873-224-147 (1-345) x AAR6351 (1-115)
 QY 1 AAGACGACACTCTTAACCAAGAAAGAAACCAACCCGCGCACAG 60
 DB 1 MetSerThrLeuProLysProGlnArgLysThrLysArgAsnThrAsnProGlyHisArg 20
 QY 61 ACGTTAAGTCCGACGCGCGGTGAGATCGTGTGAGATTAGCTGTACACGACGAG 120
 DB 21 ThrLeuSerSerGlnAlaValArgSerLeuValGlnPheThrCysTyrHisAlaIly 40
 QY 121 GCCCCAGTGGGTGCGTGCAGTGCAGCAAGACTTCCGACGCGTGCAGACCTGCGAC 180
 DB 41 AlaProSerTrpValCysValGlnCysAlaArgLeuProSerGlyArgAsnLeuAlaVal 60
 QY 181 GCGGCAACCCATCCCGACGCGCGCGGACGAGGAGAGTCTGGGCTACGCCGGGT 240
 DB 61 GlyAlaAsnProSerProGlyArgAlaGlnProArgAlaGlyProGlyLeuSerProGly 80
 QY 241 ACCCTTGGCCCTATATGGAATGAGGCGTGGGCGGAGGAGGCTCTGTCGCCGC 300
 DB 81 ThrLeuGlyProTyrMetClyMetArgAlaAlaGlyGlyGlnGlySerCysProArg 100
 QY 301 GCGGCTCTGCGCCGCTGCGGCGCCCAATGAACCCCGCGCGACGAG 345
 DB 101 AlaAlaLeuAlaArgArgGlyAlaGlnMetThrProGlyAlaGly 115
 RESULT 2
 ID AAR6550 standard; peptide: 115 AA.

XX
 AC AAR6550;
 XX
 DT 11-MAR-1997 (first entry)
 XX
 DE Hepatitis C virus type 10a isolate NE98 amino acids 1-317.
 XX
 KM Hepatitis C virus; subtype: polymerase chain reaction; amplification;
 KM PCR; primer: probe; antibody; infection.
 XX
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT MISC-difference 17
 FT /note= "amino acid residue at this position is not known"
 XX
 PN W09613590-A2.
 XX
 PD 09-MAV-1996.
 XX
 PF 23-OCT-1995; 95MO-EP04155.
 XX
 PR 28-JUN-1995; 95EP-0870076.
 PR 21-OCT-1994; 94EP-0870166.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Maertens G, Stuyver L;
 XX
 DR WPI: 1996-251460/25.
 DR N-PSDB: AAT27961.
 XX
 PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
 PT used to develop probes and primers for new sub:types and vaccines
 PT to prevent and treat infection
 XX
 PS Claim 25; Fig 3; 150pp: English.
 XX
 CC The sequences AAR6526-R96578 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-1, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the
 CC genome. This sequence represents amino acids 1-115 from the HCV type 10a
 CC isolate NE98.
 CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
 XX
 SQ Sequence 115 AA:
 Alignment Scores:
 Pred. No.: 6,25e-90 Length: 115
 Score: 98.00 Matches: 98
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 85.22% Indels: 0
 DB: Gaps: 0
 US-09-873-224-147 (1-345) x AAR96550 (1-115)
 QY 51 CGGCGACAGACGTTAAGTCCACGCGCGCGACATGCTGTCGAGATTACGTGCTA 110
 DB 18 ArgProGlnAspValLysPheProGlyGlyGlyGlnIleValGlnIleValIlyrValIleu 37
 QY 111 CCACGACGAGGCGCCCGAGTTGGGTGCTGCGAGTGCAGACAGACTTCCGAGCGGTGCGNA 170


```

Db      |||
38      ProArGATgGlyProGlnLeuGlyValAlaValAlaArgysrHsrSerLunArgSerGln 57
QY      171 CCTCCAGTAGAGCCCAACCCATCCCGAGGCGCCGAGACGAGGCGAGTCTGGGCT 230
Db      |||
58      ProArGserArGATgGlnProLleProArGAlaArGArGhrGlnGlyArGserTrpAla 77
QY      231 CAGCCCGGATACCTTGGCCCTATATAGGAATGAGGCTGGCGGGTGGGAGGCGTGC 290
Db      |||
78      GlnProGlyTrpProTrpProLeuTrpGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeu 97
QY      291 CTGTCCCGCGGCGCTCTGCGCGGTGGTGGGCGCCCAATGACCCCGCGCAGG 344
Db      |||
98      LeuSerProArGlySerArGProSerTrpGlyProAsnAspProArGArGArg 115

```

RESULT 3

AAr37931
ID AAR37931 standard; Protein; 124 AA.

AC AAR37931;

DT 23-SEP-1993 (first entry)

DE HCV core region from donor E-DL.

KX Non-coding region; hepatitis C virus; blood donor; type 2; type 1;

KM HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.

OS Hepatitis C virus.

PN WO9310239-A.

PD 27-MAY-1993.

PF 20-NOV-1992; 92WO-GB02143.

PR 21-NOV-1991; 91GB-0024696.

PR 24-JUN-1992; 92GB-0013362.

PA (COMM-) COMMON SERVICES AGENCY.

PI Chan S, Simmonds P, Yap PL;

DR WPI; 1993-182554/22.

PT DNA encoding antigenic peptide(s) of new types of hepatitis C
virus - for diagnosing and treating HCV infection, screening
PT blood samples and identifying different HCV types

PS Disclosure; Fig 7; 120pp; English.

CC This sequence shows the core region of hepatitis C virus (HCV) samples
CC from a blood donor. Analysis of this and other regions of the HCV
CC genome revealed the existence of three distinct groups of HCV.
CC Analysis of the region encompassing -255 to -62 of the 5' non coding
CC region showed a difference of 9-14% in the nucleotide sequence between
CC the three groups. Two of the groups identified were similar to those
CC of HCV variants termed type 1 and 2, whilst the third appeared to
CC represent a novel type of virus. Comparison of the NS3 region (see
CC AAR37927-30) showed a high degree of sequence diversity with type 3
CC being phylogenetically different to type 1 and 2. The same degree
CC of differ-entiation was noted in the NS-5 (see AAR37923-26) and core
CC region between type 3 and type 1 sequences.

SO Sequence 124 AA;

Alignment Scores:

Pred. No.: 1.34e-35 length: 124
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38-26% Indels: 0
DB: 14 Gaps: 0

US-09-873-224-147 (1-345) x AAR37931 (1-124)

```

QY      213 GAGGGAGAGTCTGGGCTGACCCGGGTACCTTGGCCCTATATGGAATGAGGCTGC 272
Db      |||
68      GlnGlyArGserTrpAlaGlnProGlyTyPrProTrpProLeuTrpGlyAsnGlnGlyCys 87
QY      273 GGGTGGCAGAGTGGCTGCTGTCCCGCGGCGGCTGCGCCCTGCTGGGCGCCCAATGAC 332
Db      |||
88      GLyTrpAlaGlyTrpLeuLeuSerProArGlySerArGProSerTrpAlaProAsnAsp 107
QY      333 CCGCGCGCGCAGG 344
Db      |||
108      ProArGArGArg 111

```

RESULT 4

AAr63359

ID AAR63359 standard; Protein; 166 AA.

AC AAR63359;

DT 15-AUG-1995 (first entry)

DE Hepatitis C virus core/E1 polypeptide.

KX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;

KM Classification; immunisation; prophylaxis; serotyping.

OS Hepatitis C virus (clone GB809-2).

PN WO9425601-A.

PD 10-NOV-1994.

PF 27-APR-1994; 94WO-EP01323.

PR 27-APR-1993; 93EP-0401099.

PR 05-AUG-1993; 93EP-0402019.

PA (INNO-) INNOGENETICS NV SA.

PI Maertens G, Stuyver L;

DR WPI; 1994-358277/44.

PT N-PSDB; AA078097.

PT New polynucleotide sequences from hepatitis C virus - and related
vectors, polypeptide(s) and antibodies; useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates

PS Example 10; Page 213-214; 404pp; English.

CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the core/E1 region
CC of HCV.

SO Sequence 166 AA;

Alignment Scores:

Pred. No.: 1.29e-35 length: 166
Score: 44.00 Matches: 44

Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 38.26%
 DB: 15
 Gaps: 0

US-09-873-224-147 (1-345) x AAR92968 (1-191)

QY 213 GAGGCGAGGCTCTGAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGC 272
 DB 72 GAGGCGAGGCTCTGAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGC 91
 QY 273 GGGTGGGCGAGGCTCTGAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGC 332
 DB 92 GAGGCGAGGCTCTGAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGC 111
 QY 333 CCGGCGGCGAGG 344
 DB 112 ProArgArgArg 115

RESULT 5

ID AAR92968 standard; Protein: 191 AA.

AC AAR92968;

DT 02-OCT-1996 (first entry)

DE Hepatitis C virus isolate HK10 core protein.

KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX hepatitis.

OS Hepatitis C virus.

PN W09605315-A2.

PD 22-FEB-1996.

PF 15-AUG-1995; 95WO-US10398.

PR 15-AUG-1994; 94US-0290665.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (USSH) US SEC DEPT HEALTH.

PI Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

DR N-PSDB; AAT16642.

PT DNA and amino acid sequence of HCV envelope 1 and core proteins -

PS used to determine HCV genotype and as vaccines against HCV infection.

XX Claim 4; Page 207; 340pp; English.

XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may
 CC be used in the prevention of HCV infection.

XX Sequence 191 AA;

XX SQ

Alignment Scores:
 Pred. No.: 1.26e-35 Length: 191
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conserved: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: 17 Gaps: 0

US-09-873-224-147 (1-345) x AAR92968 (1-191)

QY 213 GAGGCGAGGCTCTGAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGC 272
 DB 72 GAGGCGAGGCTCTGAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGC 91
 QY 273 GGGTGGGCGAGGCTCTGAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGC 332
 DB 92 GAGGCGAGGCTCTGAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGC 111
 QY 333 CCGGCGGCGAGG 344
 DB 112 ProArgArgArg 115

RESULT 6

ID AAR92969 standard; Protein: 191 AA.

AC AAR92969;

DT 02-OCT-1996 (first entry)

DE Hepatitis C virus isolate S52 core protein.

KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;

XX hepatitis.

OS Hepatitis C virus.

PN W09605315-A2.

PD 22-FEB-1996.

PF 15-AUG-1995; 95WO-US10398.

PR 15-AUG-1994; 94US-0290665.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

PA (USSH) US SEC DEPT HEALTH.

PI Bukh J, Miller RH, Purcell RH;

XX WPI; 1996-139709/14.

DR N-PSDB; AAT16643.

PT DNA and amino acid sequence of HCV envelope 1 and core proteins -

PS used to determine HCV genotype and as vaccines against HCV infection.

XX Claim 4; Page 208; 340pp; English.

XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may
 CC be used in the prevention of HCV infection.

XX Sequence 191 AA;

XX SQ

Alignment Scores:
 Pred. No.: 1.26e-35 Length: 191
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conserved: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: 17 Gaps: 0

US-09-873-224-147 (1-345) x AAR92969 (1-191)

QY 213 GAGGCGAGGCTCTGAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGC 272

```
|||||
Db 72 GluGIYArgSerTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGluGlyCys 91
OY 273 GGGTGGGCAAGGCTGCTGTCCCGCGGCTCTCGCCCTGTGGAGGCCCAATGAC 332
    |||||||
Db 92 GlyTrpAlaGlyTyrPneuleuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111
OY 333 CCCCCGCGCAGG 344
    |||||||
Db 112 ProArgArgArg 115

RESULT 7
AAR92970
ID AAR92970 standard; Protein; 191 AA.
XX
AC AAR92970;
XX
DT 02-OCT-1996 (first entry)
XX
DE Hepatitis C virus isolate S2 core protein.
XX
KW HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis.
XX
OS Hepatitis C virus.
XX
PN MO9605315-A2.
XX
PD 22-FEB-1996.
XX
PF 15-AUG-1995; 95WO-US10398.
XX
PR 15-AUG-1994; 94US-0290665.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH ) US SEC DEPT HEALTH.
XX
PI Bukh J, Miller RH, Purcell RH;
XX
DR WPI; 1996-139709/14.
DR N-PSDB; AAT16644.
XX
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
XX
XX
PS Claim 4; Page 209; 340pp; English.
XX
CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
CC isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may
CC be used in the prevention of HCV infection.
XX
SO Sequence 191 AA;

Alignment Scores:
Pred. No.: 1.26e-35 Length: 191
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: 17 Gaps: 0

US-09-873-224-147 (1-345) x AAR92970 (1-191)
OY 213 GAGGGAGGCTCCGTGGCTCAGCCCGGTAACCTTGGCCCTATATGGAATGAGGCTGC 272
    |||||||
Db 72 GluGIYArgSerTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGluGlyCys 91
OY 273 GGGTGGGCAAGGCTGCTGTCCCGCGGCTCTCGCCCTGTGGAGGCCCAATGAC 332
    |||||||
Db 112 ProArgArgArg 115
```

```
|||||
Db 92 GlyTrpAlaGlyTyrPneuleuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111
OY 333 CCCCCGCGCAGG 344
    |||||||
Db 112 ProArgArgArg 115

RESULT 8
AAR92971
ID AAR92971 standard; Protein; 191 AA.
XX
AC AAR92971;
XX
DT 02-OCT-1996 (first entry)
XX
DE Hepatitis C virus isolate DK12 core protein.
XX
KW HCV; EI; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis.
XX
OS Hepatitis C virus.
XX
PN MO9605315-A2.
XX
PD 22-FEB-1996.
XX
PF 15-AUG-1995; 95WO-US10398.
XX
PR 15-AUG-1994; 94US-0290665.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH ) US SEC DEPT HEALTH.
XX
PI Bukh J, Miller RH, Purcell RH;
XX
DR WPI; 1996-139709/14.
DR N-PSDB; AAT16645.
XX
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
XX
XX
PS Claim 4; Page 209-210; 340pp; English.
XX
CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
CC isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may
CC be used in the prevention of HCV infection.
XX
SO Sequence 191 AA;

Alignment Scores:
Pred. No.: 1.26e-35 Length: 191
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: 17 Gaps: 0

US-09-873-224-147 (1-345) x AAR92971 (1-191)
OY 213 GAGGGAGGCTCCGTGGCTCAGCCCGGTAACCTTGGCCCTATATGGAATGAGGCTGC 272
    |||||||
Db 72 GluGIYArgSerTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGluGlyCys 91
OY 273 GGGTGGGCAAGGCTGCTGTCCCGCGGCTCTCGCCCTGTGGAGGCCCAATGAC 332
    |||||||
Db 92 GlyTrpAlaGlyTyrPneuleuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111
OY 333 CCCCCGCGCAGG 344
```

```

Db      112 ProArgArgArg 115
RESULT 9
ID      AAR92972 standard; Protein; 191 AA.
AC      AAR92972;
XX
XX      AAR92972;
AC      AAR92972;
XX
XX      02-OCT-1996 (first entry)
DE      Hepatitis C virus isolate 24 core protein.
XX
XX      HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KM      hepatitis.
XX
XX      Hepatitis C virus.
OS      Hepatitis C virus.
PN      WO9605315-A2.
XX
XX      22-FEB-1996.
PD      15-AUG-1995; 95MO-US10398.
XX
XX      15-AUG-1994; 94US-0290665.
PR      15-AUG-1994; 94US-0290665.
XX
XX      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA      (USSH ) US SEC DEPT HEALTH.
XX
XX      Bukh J, Miller RH, Purcell RH;
PI      WPI; 1996-139709/14.
XX      N-PSDB; AAT16646.
XX
XX      DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT      used to determine HCV genotype and as vaccines against HCV infection
XX
XX      Claim 4; Page 210-211; 340pp; English.
XX
XX      AAR92936-R92987 are HCV core proteins derived from 52 different HCV
CC      isolates. Isolated cDNA sequences are used for the prodn. of primers
CC      useful for detecting the presence of HCV in a sample, the primers
CC      are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC      can be used in vaccines for immunising against HCV infection. The
CC      proteins may also be used to detect antibodies against HCV in serum,
CC      saliva, lymphocytes or other mononuclear cells. The antibodies may
CC      be used in the prevention of HCV infection.
XX
XX      SQ Sequence 191 AA;
XX
XX      Alignment Scores:
XX      Pred. No.: 1.26e-35 Length: 191
XX      Score: 44.00 Matches: 44
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 38.26% Indels: 0
XX      DB: 17 Gaps: 0
XX
XX      US-09-873-224-147 (1-345) x AAR92972 (1-191)
XX
XX      QY 213 GAGGCGAGGCTCTGGGCTCAGCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 272
XX      |||||||
XX      DB 72 GUGUyArGserTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyAsnGluGlyCys 91
XX
XX      QY 273 GGGTGGGACAGGCTCTGCTGCCCGGCGGCTCTGCGCCGCTGCGGGGCCCAATGAC 332
XX      |||||||
XX      DB 92 GLyTrpAlaGlyTrpLeuLeuSerProArGlySerArGProSerTrpGlyProAsnAsp 111
XX
XX      QY 333 CCGCGGCGCAGG 344
XX      |||||||
XX      DB 112 ProArgArgArg 115
XX
XX      RESULT 10

```

```

AAR92973
ID      AAR92973 standard; Protein; 191 AA.
XX
XX      AAR92973;
AC      AAR92973;
XX
XX      02-OCT-1996 (first entry)
DE      Hepatitis C virus isolate 28 core protein.
XX
XX      Hepatitis C virus.
OS      Hepatitis C virus.
PN      WO9605315-A2.
XX
XX      22-FEB-1996.
PD      15-AUG-1995; 95MO-US10398.
XX
XX      15-AUG-1994; 94US-0290665.
PR      15-AUG-1994; 94US-0290665.
XX
XX      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA      (USSH ) US SEC DEPT HEALTH.
XX
XX      Bukh J, Miller RH, Purcell RH;
PI      WPI; 1996-139709/14.
XX      N-PSDB; AAT16647.
XX
XX      DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT      used to determine HCV genotype and as vaccines against HCV infection
XX
XX      Claim 4; Page 211-212; 340pp; English.
XX
XX      AAR92936-R92987 are HCV core proteins derived from 52 different HCV
CC      isolates. Isolated cDNA sequences are used for the prodn. of primers
CC      useful for detecting the presence of HCV in a sample, the primers
CC      are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC      can be used in vaccines for immunising against HCV infection. The
CC      proteins may also be used to detect antibodies against HCV in serum,
CC      saliva, lymphocytes or other mononuclear cells. The antibodies may
CC      be used in the prevention of HCV infection.
XX
XX      SQ Sequence 191 AA;
XX
XX      Alignment Scores:
XX      Pred. No.: 1.26e-35 Length: 191
XX      Score: 44.00 Matches: 44
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 38.26% Indels: 0
XX      DB: 17 Gaps: 0
XX
XX      US-09-873-224-147 (1-345) x AAR92973 (1-191)
XX
XX      QY 213 GAGGCGAGGCTCTGGGCTCAGCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 272
XX      |||||||
XX      DB 72 GUGUyArGserTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyAsnGluGlyCys 91
XX
XX      QY 273 GGGTGGGACAGGCTCTGCTGCCCGGCGGCTCTGCGCCGCTGCGGGGCCCAATGAC 332
XX      |||||||
XX      DB 92 GLyTrpAlaGlyTrpLeuLeuSerProArGlySerArGProSerTrpGlyProAsnAsp 111
XX
XX      QY 333 CCGCGGCGCAGG 344
XX      |||||||
XX      DB 112 ProArgArgArg 115
XX
XX      RESULT 11
XX      AAR92974
XX      ID AAR92974 standard; Protein; 191 AA.
XX      AC AAR92974;

```

```

XX 02-OCT-1996 (first entry)
DT Hepatitis C virus isolate Z1 core protein.
XX
XX Hepatitis C virus isolate Z1 core protein.
DE
XX HCV: E1: envelope 1; core protein; HCV genotyping; antibody; vaccine;
KM hepatitis.
XX
XX Hepatitis C virus.
OS
XX WO9605315-A2.
XX
XX 22-FEB-1996.
PD
XX 15-AUG-1995; 95WO-US10398.
XX
XX 15-AUG-1994; 94US-0290665.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH ) US SEC DEPT HEALTH.
XX
XX Bukh J, Miller RH, Purcell RH;
XX
XX WPI: 1996-139709/14.
XX
XX N-PSDB; AAT16648.
XX
XX DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
XX
XX Claim 4; Page 212; 340pp; English.
XX
XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
CC isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may
CC be used in the prevention of HCV infection.
XX
XX SQ Sequence 191 AA:
XX
XX Alignment Scores:
XX Pred. No.: 1,26e-35 Length: 191
XX Score: 44.00 Matches: 44
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 38.26% Indels: 0
XX DB: 17 Gaps: 0
XX
XX US-09-873-224-147 (1-345) x AAR92974 (1-191)
XX
XX 213 GAGGGCAGGCTCTGGGCTCAGCCCGGTACCTTGCCCTATATGGGAATGAGGGCTGC 272
XX |||||||
XX Db 72 GlnblyArGserTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyAsnGlyGlyCys 91
XX |||||||
XX 273 GGGTGGGCGAGGCTGCTCTGCTCCCGCGGCTCTGCGCCGTCGTGGGCGCCAAATGAC 332
XX |||||||
XX Db 92 GlyTrpAlaGlyTrpLeuLeuSerProArGlySerArGProSerTrpGlyProAsnAsp 111
XX |||||||
XX 333 CCGCGGCGCAGG 344
XX |||||||
XX Db 112 ProArGArGArG 115
XX |||||||
XX
XX RESULT 12
XX AAR92976
XX ID AAR92976 standard; Protein: 191 AA.
XX
XX AAR92976:
XX
XX 02-OCT-1996 (first entry)
XX
XX Hepatitis C virus isolate Z6 core protein.
XX

```

```

XX HCV: E1: envelope 1; core protein; HCV genotyping; antibody; vaccine;
KM hepatitis.
XX
XX Hepatitis C virus.
OS
XX WO9605315-A2.
XX
XX 22-FEB-1996.
PD
XX 15-AUG-1995; 95WO-US10398.
XX
XX 15-AUG-1994; 94US-0290665.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH ) US SEC DEPT HEALTH.
XX
XX Bukh J, Miller RH, Purcell RH;
XX
XX WPI: 1996-139709/14.
XX
XX N-PSDB; AAT16650.
XX
XX DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
XX
XX Claim 4; Page 214; 340pp; English.
XX
XX AAR92936-R92987 are HCV core proteins derived from 52 different HCV
CC isolates. Isolated cDNA sequences are used for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may
CC be used in the prevention of HCV infection.
XX
XX SQ Sequence 191 AA:
XX
XX Alignment Scores:
XX Pred. No.: 1,26e-35 Length: 191
XX Score: 44.00 Matches: 44
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 38.26% Indels: 0
XX DB: 17 Gaps: 0
XX
XX US-09-873-224-147 (1-345) x AAR92976 (1-191)
XX
XX 213 GAGGGCAGGCTCTGGGCTCAGCCCGGTACCTTGCCCTATATGGGAATGAGGGCTGC 272
XX |||||||
XX Db 72 GlnblyArGserTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyAsnGlyGlyCys 91
XX |||||||
XX 273 GGGTGGGCGAGGCTGCTCTGCTCCCGCGGCTCTGCGCCGTCGTGGGCGCCAAATGAC 332
XX |||||||
XX Db 92 GlyTrpAlaGlyTrpLeuLeuSerProArGlySerArGProSerTrpGlyProAsnAsp 111
XX |||||||
XX 333 CCGCGGCGCAGG 344
XX |||||||
XX Db 112 ProArGArGArG 115
XX |||||||
XX
XX RESULT 13
XX AAR92977
XX ID AAR92977 standard; Protein: 191 AA.
XX
XX AAR92977:
XX
XX 02-OCT-1996 (first entry)
XX
XX Hepatitis C virus isolate Z7 core protein.
XX
XX HCV: E1: envelope 1; core protein; HCV genotyping; antibody; vaccine;
KM hepatitis.
XX

```

OS Hepatitis C virus.
 XX
 PN WO9605315-A2.
 XX
 PD 22-FEB-1996.
 XX
 PF 15-AUG-1995; 95WO-US10398.
 XX
 PR 15-AUG-1994; 94US-0290665.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.
 XX
 PI Bukh J, Miller RH, Purcell RH;
 DR WPI: 1996-139709/14.
 DR N-PSDB; AAT16651.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection
 XX
 PS Claim 4; Page 214-215; 340pp; English.
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may
 CC be used in the prevention of HCV infection.
 XX
 SQ Sequence 191 AA;
 Alignment Scores:
 Pred. No.: 1,26e-35 Length: 191
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: 17 Gaps: 0
 US-09-873-224-147 (1-345) x AAR92977 (1-191)
 QY 213 GAGGCGAGGCTCGGCTCAGCCCGGGAACCTTGCCCTATATGGGAATGAGGCTGC 272
 Db 72 GUGlyArgSerTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyAsnGlnGlyCys 91
 QY 273 GGGTGGGAGGGTGGCTGCTGCTGCCCCGGCGGCTCTGCCCGCTGCGGGGCCCAATGAC 332
 Db 92 GlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111
 QY 333 CCCCAGGCGCAGG 344
 Db 112 ProArgArgArg 115
 RESULT 14
 AAR92978
 ID AAR92978 standard; Protein: 191 AA.
 XX
 AC AAR92978;
 XX
 DT 02-OCT-1996 (first entry)
 XX
 DE Hepatitis C virus isolate DK13 core protein.
 XX
 KW HCV; E1: envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9605315-A2.
 XX

PD 22-FEB-1996.
 XX
 PF 15-AUG-1995; 95WO-US10398.
 XX
 PR 15-AUG-1994; 94US-0290665.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.
 XX
 PI Bukh J, Miller RH, Purcell RH;
 DR WPI: 1996-139709/14.
 DR N-PSDB; AAT16652.
 XX
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
 PT used to determine HCV genotype and as vaccines against HCV infection
 XX
 PS Claim 4; Page 215-216; 340pp; English.
 CC AAR92936-R92987 are HCV core proteins derived from 52 different HCV
 CC isolates. Isolated cDNA sequences are used for the prodn. of primers
 CC useful for detecting the presence of HCV in a sample, the primers
 CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
 CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may
 CC be used in the prevention of HCV infection.
 XX
 SQ Sequence 191 AA;
 Alignment Scores:
 Pred. No.: 1,26e-35 Length: 191
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: 17 Gaps: 0
 US-09-873-224-147 (1-345) x AAR92978 (1-191)
 QY 213 GAGGCGAGGCTCGGCTCAGCCCGGGAACCTTGCCCTATATGGGAATGAGGCTGC 272
 Db 72 GUGlyArgSerTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyAsnGlnGlyCys 91
 QY 273 GGGTGGGAGGGTGGCTGCTGCTGCCCCGGCGGCTCTGCCCGCTGCGGGGCCCAATGAC 332
 Db 92 GlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111
 QY 333 CCCCAGGCGCAGG 344
 Db 112 ProArgArgArg 115
 RESULT 15
 AAR67589
 ID AAR67589 standard; Protein: 502 AA.
 XX
 AC AAR67589;
 XX
 DT 07-SEP-1995 (first entry)
 XX
 DE Hepatitis C virus YS117-5' gene product.
 XX
 KW Hepatitis C virus; HCV; non-A non-B; YS117-5'; treatment.
 XX
 OS Hepatitis C virus.
 XX
 PN JP06319563-A.
 XX
 PD 22-NOV-1994.
 XX
 PF 13-MAY-1993; 93JP-0147133.
 XX
 PR 13-MAY-1993; 93JP-0147133.
 XX

XX (TMO) IMMUNO JAPAN KK.
PA WPI, 1995-040318/06.
XX N-PSDB; AAO79141.
DR A hepatitis C virus gene and oligo-nucleotide(s) - used for the
XX treatment of hepatitis C
PT
XX
PS Claim 19: Page 34, 41pp; Japanese.
XX
CC AAO79141 is the hepatitis C virus (HCV) gene YS117-3' cDNA, it encodes
CC the protein described in AAR67588. Both the cDNA and protein can be
CC used in the treatment of HCV infection
XX

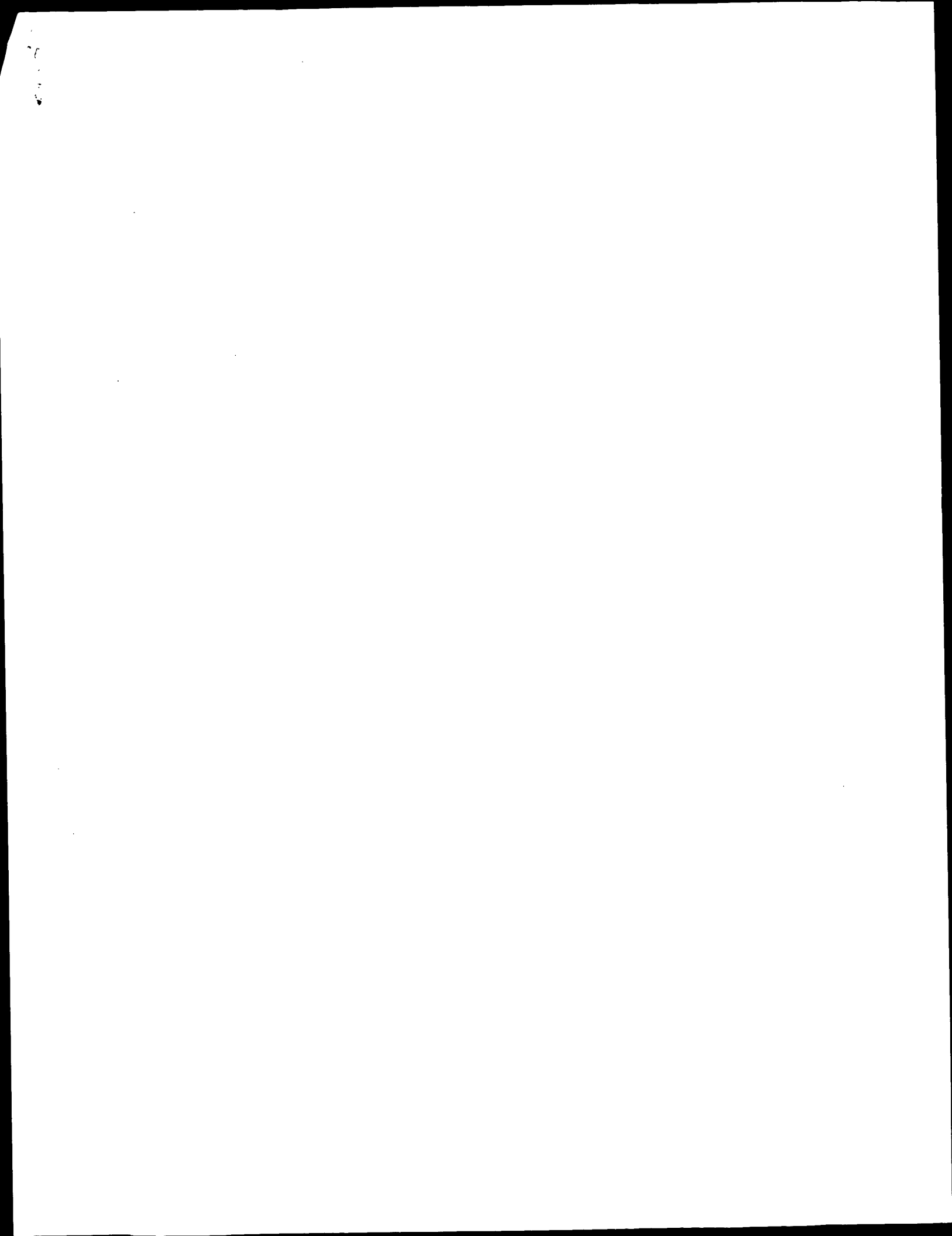
SO Sequence 502 AA:

Alignment Scores:
Pred. No.: 1.19e-29 Length: 502
Score: 38.00 Matches: 38
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 33.04% Indels: 0
DB: 16 Gaps: 0

US-09-873-224-147 (1-345) x AAR67589 (1-502)

OY 213 GAGGCGAGGCTCCTGGGCTCAGCCCGGGTACCCCTTGCCCTATATGGAATGAGGGCTGC 272
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
72 GIUGLYARGSerTPpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGluGlyCys 91
OY 273 GAGTGGGCGAGGCTGCTGCTGCTCCCGCGGGCTCTGCGCCCGTGGGGGCCA 326
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
92 GlyTPpAlaGlyTyrPLeuLeuSerProArgGlySerArgProSerTrpGlyPro 109

Search completed: February 19, 2003, 02:24:20
Job time : 36 secs



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 19, 2003, 02:17:50 ; Search time 17.5 Seconds
(without alignments)
3790.443 Million cell updates/sec

Title: US-09-873-224-147

Sequence: 1 ATGAGACACTTCTTAACC.....AAATGACCCCGCGCAGGA 345

Scoring table: OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565536

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+np.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09873224/runat_14022003_092759_28172/app_query.fasta.1.519
-DB=pir.73 -OPMT=fastan -SUFFIX=olin2p.rpr -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09873224_@CGN_1_1_38_@runat_14022003_092759_28172 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMOUT=120
-WARN_TIMOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database: PIR.73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	91	79.1	108	S44190	genome polyprotein
2	44	38.3	115	S41364	genome polyprotein
3	44	38.3	266	P00393	genome polyprotein
4	43	37.4	114	S41365	genome polyprotein
5	43	37.4	114	S41359	genome polyprotein
6	43	37.4	114	S41362	genome polyprotein
7	43	37.4	114	S41358	genome polyprotein
8	43	37.4	114	S41366	genome polyprotein
9	43	37.4	124	S41360	genome polyprotein
10	43	37.4	411	PC2060	genome polyprotein
11	43	37.4	411	PC2061	genome polyprotein
12	37	32.2	492	S41288	genome polyprotein
13	34	29.6	112	S41341	genome polyprotein
14	34	29.6	115	S41344	genome polyprotein

15	34	29.6	115	2	S41345	genome polyprotein
16	34	29.6	118	2	S41346	genome polyprotein
17	34	29.6	123	2	S41361	genome polyprotein
18	34	29.6	178	2	PS0388	genome polyprotein
19	34	29.6	189	2	S32740	polyprotein - hepa
20	34	29.6	322	2	JN0265	genome polyprotein
21	34	29.6	640	2	J01584	genome polyprotein
22	34	29.6	3011	1	GNMWC3	genome polyprotein
23	34	29.6	3011	1	GNMWC4	genome polyprotein
24	34	29.6	3011	1	S40770	structural protein
25	29	25.2	513	2	A44150	genome polyprotein
26	27	23.5	108	2	S41356	genome polyprotein
27	27	23.5	874	2	J00883	genome polyprotein
28	27	23.5	3033	1	GNMWJ8	genome polyprotein
29	25	21.7	3010	1	A45573	genome polyprotein
30	24	20.9	441	2	S12707	genome polyprotein
31	23	20.0	117	2	S41363	genome polyprotein
32	22	19.1	874	2	J00881	genome polyprotein
33	22	19.1	3033	1	J01303	core protein
34	18	15.7	82	2	P0804	genome polyprotein
35	18	15.7	82	2	S21336	genome polyprotein
36	18	15.7	108	2	S41353	genome polyprotein
37	18	15.7	108	2	S41357	genome polyprotein
38	18	15.7	108	2	S41357	genome polyprotein
39	18	15.7	108	2	S41348	genome polyprotein
40	18	15.7	112	2	S41371	genome polyprotein
41	18	15.7	114	2	S41370	genome polyprotein
42	18	15.7	114	2	S41369	genome polyprotein
43	18	15.7	114	2	S41368	genome polyprotein
44	18	15.7	115	2	S41342	genome polyprotein
45	18	15.7	115	2	S41350	genome polyprotein

ALIGNMENTS

RESULT 1
S44190
genome polyprotein - hepatitis C virus (genotype 3, N6) (fragment)
N:Contains: core protein
C:Species: hepatitis C virus
A:Variety: genotype 3, N6
C:Date: 07-Oct-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
C:Accession: S44190
R:van Doorn, L.J.; Kleier, G.F.M.; Brouwer, J.F.
submitted to the EMBL Data Library, April 1994
A:Description: Analysis of Hepatitis C virus genotypes 1 to 5.
A:Reference number: S44190
A:Accession: S44190
A:Molecule type: genomic RNA
A:Residues: 1-108 <VAN>
A:Cross-references: EMBL:X78863; NID:g474329; PIDN:CA55412.1; PID:g474330
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: capsid protein; core protein; polyprotein
F:1-108/Product: core protein #status predicted <MAT>

Alignment Scores:

Pred. No.: 1.32e-88 Length: 108
Score: 91.00 Matches: 91
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 79.13% Indels: 0
DB: Gaps: 0

US-09-873-224-147 (1-345) x S44190 (1-108)

QY	51	CGGCCACAGAGCGTTAAGTTCACGCGCGCTCAGATCGTTGGTGAAGTTTACGTGCTA	110
DB	18	ArpProGlnAspValIysPheProGlyGlnIleValIeIyGlyValIyrrValIeu	37
QY	111	CCAGCAGAGGCGCCCACTTGGTGGTGGCGAGTGGCAGACTCCGAGCGGTGGCAA	170
DB	38	ProArGArGlyProGlnLeuGlyValArGAlaValArGlySthrSerGluArGserGln	57

Keywords: capsid protein

[illegible]

RESULT: 5
341359

genome polyprotein - hepatitis C virus (genotype 3, N2) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 3, N2
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41359
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 Submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41359
 A:Molecule type: genomic RNA
 A:Residues: 1-114 <VAN>
 A:Cross-references: EMBL:Z29462; NID:g443886; PIDN:CAA82600.1; PID:g443887
 A:Experimental source: genotype 3, N2
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; polypeptide
 F:1-114/Product: core protein #status predicted <MAT>

Alignment Scores:

Pred. No.:	3.01e-37	Length:	114
Score:	43.00	Matches:	43
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.39%	Indels:	0
DB:	2	Gaps:	0

US-09-873-224-147 (1-345) x S41359 (1-114)

OY 213 GAGGCGAGGCTCGGCTGACGCCGGGTACCCCTTGCGCCCTATATGGAATGAGGCTGC 272
 |||||||
 Db 72 GlnGlyArgSerTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyLysnGlnGlyCys 91
 OY 273 GGGTGGCAGAGGCTGCTCTGCTCCCGCGGCTCTCGCCGCTGCGGCGCCCAATGAC 332
 |||||||
 Db 92 GlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111
 OY 333 CCCCCGGCG 341
 |||||||
 Db 112 ProArgArg 114

RESULT 6

S41362 genome polypeptide - hepatitis C virus (genotype 4, N1) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 4, N1
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41362
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 Submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41362
 A:Molecule type: genomic RNA
 A:Residues: 1-114 <VAN>
 A:Cross-references: EMBL:Z29465; NID:g443892; PIDN:CAA82603.1; PID:g443893
 A:Experimental source: genotype 4, N1
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; polypeptide
 F:1-114/Product: core protein #status predicted <MAT>

Alignment Scores:

Pred. No.:	3.01e-37	Length:	114
Score:	43.00	Matches:	43
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.39%	Indels:	0
DB:	2	Gaps:	0

US-09-873-224-147 (1-345) x S41362 (1-114)

OY 213 GAGGCGAGGCTCGGCTGACGCCGGGTACCCCTTGCGCCCTATATGGAATGAGGCTGC 272
 |||||||

Db 72 GlnGlyArgSerTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyLysnGlnGlyCys 91

OY 273 GGGTGGCAGAGGCTGCTCTGCTCCCGCGGCTCTCGCCGCTGCGGCGCCCAATGAC 332
 |||||||

Db 92 GlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111

OY 333 CCCCCGGCG 341
 |||||||
 Db 112 ProArgArg 114

RESULT 7

S41358 genome polypeptide - hepatitis C virus (genotype 3, N1) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 3, N1
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41358
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 Submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41358
 A:Molecule type: genomic RNA
 A:Residues: 1-114 <VAN>
 A:Cross-references: EMBL:Z29461; NID:g443884; PIDN:CAA82599.1; PID:g443885
 A:Experimental source: genotype 3, N1
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; polypeptide
 F:1-114/Product: core protein #status predicted <MAT>

Alignment Scores:

Pred. No.:	3.01e-37	Length:	114
Score:	43.00	Matches:	43
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.39%	Indels:	0
DB:	2	Gaps:	0

US-09-873-224-147 (1-345) x S41358 (1-114)

OY 213 GAGGCGAGGCTCGGCTGACGCCGGGTACCCCTTGCGCCCTATATGGAATGAGGCTGC 272
 |||||||
 Db 72 GlnGlyArgSerTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyLysnGlnGlyCys 91
 OY 273 GGGTGGCAGAGGCTGCTCTGCTCCCGCGGCTCTCGCCGCTGCGGCGCCCAATGAC 332
 |||||||
 Db 92 GlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111
 OY 333 CCCCCGGCG 341
 |||||||
 Db 112 ProArgArg 114

RESULT 8

S41366 genome polypeptide - hepatitis C virus (genotype 4, N5) (fragment)

N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 4, N5
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41366
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 Submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41366
 A:Molecule type: genomic RNA
 A:Residues: 1-114 <VAN>
 A:Cross-references: EMBL:Z29469; NID:g443900; PIDN:CAA82607.1; PID:g443901
 A:Experimental source: genotype 4, N5
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; polypeptide
 F:1-114/Product: core protein #status predicted <MAT>

276 TGGGCAAGGGTGCCTCCCTGCCCCGCGGAGCTGTGCCCCGTGCTGGGGGCCCAATGACCCC 335
|||||
93 TriplagilytripeuenseProargglyserArgProserTirpglyProasnapro 112
|||||

OY 336 CGGCGCAGG 344
 |||||||
 Db 113 ArgArgArg 115

RESULT 12

S41288
 genome polypeptide - hepatitis C virus (fragment)
 N:Contains: core protein; envelope protein; NS1 protein
 C:Species: hepatitis C virus
 C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41288
 R:Seeliger, R.
 submitted to the EMBL Data Library, December 1993
 A:Reference number: S41288
 A:Accession: S41288
 A:Molecule type: genomic RNA
 A:Residues: 1-492 <SEE>
 A:Cross-references: EMBL:X76918
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; envelope protein; nonstructural protein; poly
 F:1-191/Product: core protein #status predicted <COR>
 F:192-372/Product: envelope protein #status predicted <ENV>
 F:373-492/Product: NS1 protein (fragment) #status predicted <NS1>

Alignment Scores:

Pred. No.: 7.03e-31 Length: 492
 Score: 37.00 Matches: 37
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 32.17% Indels: 0
 DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x S41288 (1-492)

OY 213 GAGGCGAGTCTGGGCTCAGCCCGGTACCTTGCCCTATATGGAATGAGGCTGC 272
 |||||||
 Db 72 GLUGLYARGSERTRPAGLNPGLTYRPTOTRPROLEUTYRGLYSNGUGLYCYS 91

OY 273 GGGGGGAGGCTGCTCTCCCGCGGCTCCGCCCTCGGGGC 323
 |||||||
 Db 92 GLYTRPAGLTYRTRPLeuSerProArgGlySerArgProSerTrpGly 108

RESULT 13

S41341
 genome polypeptide - hepatitis C virus (genotype 1, N1) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 1, N1
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41341
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41341
 A:Molecule type: genomic RNA
 A:Residues: 1-112 <VAN>
 A:Cross-references: EMBL:Z29444; NID:q443850; PIDN:CAA82582.1; PID:q443851
 A:Experimental source: genotypel, N1
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; polypeptide
 F:1-112/Product: core protein #status predicted <MAT>

Alignment Scores:

Pred. No.: 1.28e-27 Length: 112
 Score: 34.00 Matches: 34
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 29.57% Indels: 0
 DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x S41341 (1-112)

OY 225 TGGGCTACCCGGGTACCTTGCCCTATATGGAATGAGGCTGCGGCGCAGG 284
 |||||||
 Db 76 TRPALAGLNPGLTYRPTOTRPROLEUTYRGLYSNGUGLYCYSGLYTRPALAGLY 95

OY 285 TGGCTCTGTCCCGCGGCTCGCCCGTCTGGGCGCCA 326
 |||||||
 Db 96 TRPLeuSerProArgGlySerArgProSerTrpGlyPro 109

RESULT 14

S41344
 genome polypeptide - hepatitis C virus (genotype 1, N2) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 1, N2
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41344
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41344
 A:Molecule type: genomic RNA
 A:Residues: 1-115 <VAN>
 A:Cross-references: EMBL:Z29447; NID:q443856; PIDN:CAA82585.1; PID:q443857
 A:Experimental source: genotype 1, N2
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; polypeptide
 F:1-115/Product: core protein #status predicted <MAT>

Alignment Scores:

Pred. No.: 1.28e-27 Length: 115
 Score: 34.00 Matches: 34
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 29.57% Indels: 0
 DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x S41344 (1-115)

OY 225 TGGGCTACCCGGGTACCTTGCCCTATATGGAATGAGGCTGCGGCGCAGG 284
 |||||||
 Db 76 TRPALAGLNPGLTYRPTOTRPROLEUTYRGLYSNGUGLYCYSGLYTRPALAGLY 95

OY 285 TGGCTCTGTCCCGCGGCTCGCCCGTCTGGGCGCCA 326
 |||||||
 Db 96 TRPLeuSerProArgGlySerArgProSerTrpGlyPro 109

RESULT 15

S41345
 genome polypeptide - hepatitis C virus (genotype 1, N3) (fragment)
 N:Contains: core protein
 C:Species: hepatitis C virus
 A:Variety: genotype 1, N3
 C:Date: 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 17-Nov-2000
 C:Accession: S41345
 R:van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 submitted to the EMBL Data Library, January 1994
 A:Description: Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 A:Reference number: S41341
 A:Accession: S41345
 A:Molecule type: genomic RNA
 A:Residues: 1-115 <VAN>
 A:Cross-references: EMBL:Z29448; NID:q443858; PIDN:CAA82586.1; PID:q443859
 A:Experimental source: genotype 1, N3
 C:Superfamily: hepatitis C virus genome polypeptide
 C:Keywords: capsid protein; core protein; polypeptide
 F:1-115/Product: core protein #status predicted <MAT>

Alignment Scores:

Pred. No.: 1.28e-27 Length: 115
 Score: 34.00 Matches: 34
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00%
 Query Match: 29.57%
 DB: 2
 Mismatches: 0
 Indels: 0
 Gaps: 0

US-09-873-224-147 (1-345) x S41345 (1-115)

QY	225	TGGGCTCAGCCCGGTACCCCTTGGCCCTATATGGGATAGAGGCTGCGGTGGCAGGG	284
Db	76	TriPalaglinProglyTyrProtrProleuTYRGLYAsnGLUGLYCysGLYTriPalagly	95
QY	285	TGGCTCCTGTCCCGCGGCGCTCTCGCCCGTGTGGGGCCCA	326
Db	96	TriLeuLeuSerProArgGlySerArgProSerTriPGLYPro	109

Search completed: February 19, 2003, 02:26:53
 Job time : 18.5 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: February 19, 2003, 02:15:45 ; Search time 11 seconds

(without alignments)
2601.697 Million cell updates/sec

Title: US-09-873-224-147

Sequence: 1 ATGAGCACACTCTCTAAC.....AAATGACCCCGCGCAGGA 345

Scoring table:

OLIGO	
Xgapop 60.0	Xgapext 60.0
Ygapop 60.0	Ygapext 60.0
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 112892 seqs, 41476328 residues

Word size: 1

Total number of hits satisfying chosen parameters: 225614

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cn2.1/USPTO_SPOOL/US09873224/runat_14022003_092759_28147/app-query.fasta.1.519
-DB=SwissProt_40 -OPMT=fastan -SUFFIX=olin2p.rsp -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR.SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09873224 -GCN 1.1.16 -runat 14022003_092759_28147 -NCPU=6 -ICPU=3
-NO_XLPRX -NO_MMAP -LARGEDUQUET -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	29.6	3011	1	POLG_HCV1
2	34	29.6	3011	1	POLG_HCV1
3	27	23.5	737	1	POLG_HCV7
4	27	23.5	737	1	POLG_HCV7
5	25	21.7	3033	1	POLG_HCV8
6	25	21.7	3033	1	POLG_HCV8
7	22	19.1	737	1	POLG_HCV5
8	22	19.1	737	1	POLG_HCV5
9	18	15.7	513	1	POLG_HCV2
10	18	15.7	513	1	POLG_HCV2
11	18	15.7	520	1	POLG_HCV4
12	18	15.7	520	1	POLG_HCV4
13	18	15.7	3010	1	POLG_HCVB
14	18	15.7	3010	1	POLG_HCVB
15	18	15.7	3010	1	POLG_HCVB
16	18	15.7	3010	1	POLG_HCVB
17	8	7.0	451	1	HEMN_PARDE
18	8	7.0	451	1	HEMN_PARDE
19	8	7.0	466	1	YEL4_SCHPO
20	8	7.0	466	1	YEL4_SCHPO
21	8	7.0	855	1	POLG_HRV3
22	8	7.0	855	1	POLG_HRV3
23	8	7.0	982	1	RPO_GGNV
24	8	7.0	982	1	RPO_GGNV
25	8	7.0	982	1	RPO_GGNV
26	8	7.0	982	1	RPO_GGNV
27	8	7.0	982	1	RPO_GGNV
28	8	7.0	982	1	RPO_GGNV
29	8	7.0	982	1	RPO_GGNV
30	8	7.0	982	1	RPO_GGNV
31	8	7.0	982	1	RPO_GGNV
32	8	7.0	982	1	RPO_GGNV
33	8	7.0	982	1	RPO_GGNV
34	8	7.0	982	1	RPO_GGNV
35	8	7.0	982	1	RPO_GGNV
36	8	7.0	982	1	RPO_GGNV
37	8	7.0	982	1	RPO_GGNV
38	8	7.0	982	1	RPO_GGNV
39	8	7.0	982	1	RPO_GGNV
40	8	7.0	982	1	RPO_GGNV
41	8	7.0	982	1	RPO_GGNV
42	8	7.0	982	1	RPO_GGNV
43	8	7.0	982	1	RPO_GGNV
44	8	7.0	982	1	RPO_GGNV
45	8	7.0	982	1	RPO_GGNV

C 19	7	6.1	81	1	MCPI_HIRME
C 20	7	6.1	151	1	GDPI_MOUSE
C 21	7	6.1	153	1	RIBI_RHME
C 22	7	6.1	195	1	RERB_ARATH
C 23	7	6.1	196	1	LPAC_MYCTU
C 24	7	6.1	206	1	PTCA_HUMAN
C 25	7	6.1	211	1	SGAR_HYME
C 26	7	6.1	231	1	PDX3_SCHPO
C 27	7	6.1	241	1	CS31_ECOLI
C 28	7	6.1	270	1	MER1_YEAST
C 29	7	6.1	286	1	GUB_RHOMR
C 30	7	6.1	310	1	Y497_MYCTU
C 31	7	6.1	312	1	PDX3_SCHPO
C 32	7	6.1	315	1	SORC_KLEPN
C 33	7	6.1	324	1	D3H1_DROME
C 34	7	6.1	332	1	G3P_STRAU
C 35	7	6.1	340	1	Y480_MYCTU
C 36	7	6.1	357	1	NDP1_CHICK
C 37	7	6.1	363	1	SSR5_RAT
C 38	7	6.1	371	1	OPSL_CALVI
C 39	7	6.1	371	1	V2R_HUMAN
C 40	7	6.1	372	1	GDPI_HUMAN
C 41	7	6.1	373	1	OPSL_DROME
C 42	7	6.1	374	1	OPSL_DROPS
C 43	7	6.1	379	1	YFBE_ECOLI
C 44	7	6.1	380	1	REON_CHICK
C 45	7	6.1	389	1	RIR2_HUMAN

ALIGNMENTS

RESULT 1

POLG_HCV1 STANDARD; PRT: 3011 AA.

AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitis C virus) (EC 3.4.21.99); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
DE Hepatitis C virus (isolate 1) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;
OC Hepatitis C virus.
OC NCBI_TaxID=11104;
RN [1]
RP MEDLINE-91172826; PubMed-1848704;
RX Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.;
RA "Genetic organization and diversity of the hepatitis C virus."; Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate + (RNA)(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

position, Cys or Thr in P1 and Ser or Ala in P1'.

-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA}(N).

-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1 AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

-1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.

-1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL; M67463; AAA45534.1; -.

CC PIR; A36814; GNMVCH.

CC PDB; 1HEI; 25-NOV-98.

CC PDB; 1AIV; 16-FEB-99.

CC MEROPS; S29.001; -.

CC DR MEROPS; U39.001; -.

CC TRANSFAC; T04155; -.

CC InterPro: IPR001410; DEAD.

CC InterPro: IPR002531; HCV_NS1.

CC InterPro: IPR002518; HCV_NS2.

CC InterPro: IPR004109; HCV_NS3.

CC InterPro: IPR000745; HCV_NS4.

CC InterPro: IPR001490; HCV_NS4b.

CC InterPro: IPR002868; HCV_NS5a.

CC InterPro: IPR002166; HCV_NS5b.

CC InterPro: IPR002522; HCV_capsid.

CC InterPro: IPR002521; HCV_core.

CC InterPro: IPR002519; HCV_env.

CC InterPro: IPR001650; Helicase_C.

CC Pfam; PF00271; helicase_C; 1.

CC Pfam; PF00998; HCV_RDRP; 1.

CC Pfam; PF01001; HCV_NS4b; 1.

CC Pfam; PF01006; HCV_NS4a; 1.

CC Pfam; PF01506; HCV_NS5a; 1.

CC Pfam; PF01538; HCV_NS2; 1.

CC Pfam; PF01539; HCV_env; 1.

CC Pfam; PF01542; HCV_core; 1.

CC Pfam; PF01543; HCV_capsid; 1.

CC Pfam; PF01560; HCV_NS1; 1.

CC Pfam; PF02907; HCV_NS3; 1.

CC ProDom; PD186062; HCV_NS1; 1.

CC PolyProtein; Glycoprotein; Transferrase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolyase; Serine protease; 3D-structure.

CC INIT_MET 1 1

CC CHAIN 1 191

CC FT 192 383 CAPSID PROTEIN C.

CC FT 384 746 ENVELOPE GLYCOPROTEIN E1.

CC FT 747 809 ENVELOPE GLYCOPROTEIN E2.

CC FT 809 810 PROTEIN P7.

CC FT 1026 1026 NONSTRUCTURAL PROTEIN NS2.

CC FT 1027 1657 PROTEASE/HELICASE NS3.

CC FT 1658 1711 NONSTRUCTURAL PROTEIN NS4A.

CC FT 1712 1972 NONSTRUCTURAL PROTEIN NS4B.

CC FT 1973 2420 NONSTRUCTURAL PROTEIN NS5A.

CC FT 2421 3011 NONSTRUCTURAL PROTEIN NS5B.

CC FT 3011 369 POTENTIAL.

CC FT 369 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

CC FT 1083 1107 ACT_SITE

CC FT 1107 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

CC FT 1165 1230 ACT_SITE

CC FT 1230 1319 ATP (POTENTIAL).

CC FT 1319 1316 DECH_BOX.

CC FT 1316 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC CARBOHYD 196

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 3011 AA: 327142 MW: 772CBH29CCD94753 CRR64;

Alignment Scores:

Pred. No: 6.87e-27 Length: 3011

Score: 34.00 Matches: 34

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 29.57% Indels: 0

DB: 1 Gaps: 0

US-09-873-224-147 (1-345) x POLG_HCVH (1-3011)

Qy 225 TGGGCTGACCCGGGTACCCCTTGCCCTATATGGAGATGAGGCGGTGGCGAGG 284

Db 76 TrrpalaGinProGlyIrrProIrrProenylGlynsnGlyGlyIrrpalaGly 95

Qy 285 TGGCTCTGTCCCGCGGGGCTCTGCGCCGTGCGAGGCCCA 326

Db 96 TrpleuSeuSerProArGlySerArGProSerTrpIrrpIrrp 109

RESULT 3

ID POLG_HCVJ7 STANDARD: PRI: 737 AA.

AC P27961;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix protein (Envelope protein M); Major envelope protein E; Nonstructural proteins NS1 and NS2] (Fragment).

DE Hepatitis C virus (isolate HC-J7) (HCV).

OS Vituases; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; OC Hepacivirus.

OX NCB1_TaxID=11114;

ON [1]

RX MEDLINE=92230232; PubMed=1314459;

RA Okamoto H., Kuri K., Okada S.T., Yamamoto K., Lizuka H., Tanaka T., Fukuda S., Tsuda F., Mishio S.;

RT "Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes.";

RT Virology 188:331-341(1992).

RL -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION. NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC


```

FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642BB CRC64;

Alignment Scores:
Pred. No.: 1.28e-17 Length: 3010
Score: 25.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.74% Indels: 0
DB: 1 Gaps: 0

US-09-873-224-147 (1-345) x POLG_HCVJTT (1-3010)
OY 51 CGGCCACAGACGTTAACTTCCACAGCGCGGTACGATCGTTGGAGGATTACGTCGTA 110
DB 18 AAGPProGlnAAAPValIlySpheProGlyIlyGlyIlyValIlyGlyValIlyValIleu 37
OY 111 CCACGACAGGGGCCCC 125
DB 38 ProArgArgGlyPro 42

RESULT 6
POLG_HCVJ5 STANDARD: PRT; 737 AA.
ID POLG_HCVJ5
AC P27960;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1 and NS2] (Fragment).
OS Hepatitis C virus (isolate HC-95) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11112;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.I., Yamamoto K., Litzuka H.,
RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes".
RL Virology 186:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC	EMBL	D10075	BAA00969.1	-	CC
CC	or send an email to	license@ib-sib.ch			
DR	InterPro	IPR002531	HCV_NS1		
DR	InterPro	IPR002532	HCV_capsid		
DR	InterPro	IPR002531	HCV_core		
DR	InterPro	IPR002519	HCV_env		
DR	Pfam	PF01539	HCV_env	1	
DR	Pfam	PF01542	HCV_core	1	
DR	Pfam	PF01543	HCV_capsid	1	
DR	Pfam	PF01560	HCV_NS1	1	
DR	ProDom	PD186062	HCV_NS1	1	
KW	Polypeptide		Glycoprotein		
KW	Transmembrane		Nonstructural protein		
FT	INT_MET	1			
FT	CHAIN	1	115		
FT	CHAIN	116	191		
FT	CHAIN	192	383		
FT	CHAIN	384	733		
FT	CHAIN	734	>737		
FT	TRANSMEM	347	369		
FT	CARBOHYD	196	196		
FT	CARBOHYD	209	209		
FT	CARBOHYD	234	234		
FT	CARBOHYD	305	305		
FT	CARBOHYD	417	417		
FT	CARBOHYD	423	423		
FT	CARBOHYD	430	430		
FT	CARBOHYD	448	448		
FT	CARBOHYD	477	477		
FT	CARBOHYD	534	534		
FT	CARBOHYD	542	542		
FT	CARBOHYD	558	558		
FT	CARBOHYD	578	578		
FT	CARBOHYD	627	627		
FT	CARBOHYD	649	649		
FT	NON_TER	737	737		
SQ	SEQUENCE	737 AA	81207 MM	3AF699D82AD501B1 CRC64	
Alignment Scores:					
Pred. No.:	1.85e-14	Length:	737		
Score:	22.00	Matches:	22		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	19.13%	Indels:	0		
DB:	1	Gaps:	0		
US-09-873-224-147 (1-345) x POLG_HCV95 (1-737)					
QY	273	GGGTGGGAGGAGGTGGCTCTCCCGCGGCGCTCCGCGTGGGGCCCAATGAC	332		
DB	92	GLYTPPALAGLYTTPLEULESERPROIRGILSETIRPPOSERIRPOLYPROXNASNP	111		
QY	333	CCCCCG 338			
DB	112	PROARG 113			
RESULT 7					
ID	POLG_HCV96	STANDARD:	PRT:	3033 AA	
AC	P26660				
DT	01-AUG-1992	(Rel. 23, Created)			
DT	01-AUG-1992	(Rel. 23, Last sequence update)			
DE	Genome polypeptide (Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP66) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).				

OS Hepatitis C virus (isolate HC-J6) (HCV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 OX NCBI_TaxID=11113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92044440; PubMed=1658196;
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kural K., Iizuka H.,
 RA Machida A., Miyakawa Y., Mayumi M.;
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
 RT from a human carrier: comparison with reported isolates for conserved
 RT and divergent regions.";
 RL J. Gen. Virol. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC -1- CATALYTIC ACTIVITY: A POSSIBLE ROLE IN THE VIRAL RNA REPLICATION.
 CC CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 CC {RNA}(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D00944; BA00792.1; -;
 DR PIR: J01303; J01303.
 DR HSSP: P27958; 1HBI.
 DR MEROPS: S29.001; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4A.
 DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_NS5B.
 DR InterPro: IPR002522; HCV_NS4B.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; HCV_RdRp; 1.
 DR Pfam: PF01001; HCV_NS4B; 1.
 DR Pfam: PF01006; HCV_NS4A; 1.
 DR Pfam: PF01506; HCV_NS5A; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PD186062; HCV_NSI; 1.
 KW Polypeptide; glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Serine protease.
 FT INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 733
 CHAIN 115
 CAPSID PROTEIN C (POTENTIAL).
 MATRIX PROTEIN E (POTENTIAL).
 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 NONSTRUCTURAL PROTEIN NSI (POTENTIAL).

FT CHAIN 734 1010
 FT CHAIN 1011 1619
 FT CHAIN 1620 1866
 FT CHAIN 1867 2017
 FT CHAIN 2018 3033
 FT TRANSMEM 347 363
 FT ACT_SITE 1087 1087
 FT ACT_SITE 1111 1111
 FT ACT_SITE 1169 1169
 FT NE_BIND 1234 1241
 FT SITE 1320 1323
 FT CARBOHYD 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 417 417
 FT CARBOHYD 423 423
 FT CARBOHYD 430 430
 FT CARBOHYD 448 448
 FT CARBOHYD 477 477
 FT CARBOHYD 534 534
 FT CARBOHYD 542 542
 FT CARBOHYD 558 558
 FT CARBOHYD 578 578
 FT CARBOHYD 627 627
 FT CARBOHYD 649 649
 FT CARBOHYD 1091 1091
 FT CARBOHYD 2038 2038
 FT CARBOHYD 2811 2811
 SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
 Alignment Scores:
 Pred. No.: 1,57e-14
 Score: 22.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 19.13%
 DB: 1
 Gaps: 0

US-09-873-224-147 (1-345) x POLG_HCVJ6 (1-3033)
 QY 273 GGGTGGGAGGAGGCTCTGCTCCCGCGGCTCTGCGCGGCGGAGGCGCAATGAC 332
 Db 92 G1YTPALAG1YTPLeuLeuSerProArgGlySerArgProSerIrrpGlyProAsnasp 111
 QY 333 CCCCCG 338
 Db 112 ProArg 113

RESULT 8
 POLG_HCVJ2 STANDARD; PRT; 513 AA.
 AC P27959;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus (isolate HC-J2) (HCV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 OX NCBI_TaxID=11111;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230232; PubMed=1314459;
 RA Okamoto H., Kural K., Okada S.-I., Yamamoto K., Iizuka H.,
 RA Tanaka T., Fukuda S., Tsuda F., Mishiro S.;
 RT "Full-length sequence of a hepatitis C virus genome having poor
 RT homology to reported isolates: comparative study of four distinct
 RT genotypes.";
 RL Virology 188:331-341(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: OF
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MNNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10074; BAA00968.1; -
CC InterPro: IPR002531; HCV_NSI.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC Pfam: PF01539; HCV_core; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01560; HCV_NSI; 1.
CC ProDom: PD186062; HCV_NSI; 1.
CC Polyprotein: Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC INIT_MET 1
CC 1
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 >513
CC TRANSMEM 347 369
CC CARBOHYD 196 196
CC CARBOHYD 209 209
CC CARBOHYD 233 233
CC CARBOHYD 234 234
CC CARBOHYD 250 250
CC CARBOHYD 305 305
CC CARBOHYD 417 417
CC CARBOHYD 423 423
CC CARBOHYD 430 430
CC CARBOHYD 448 448
CC CARBOHYD 513 513
CC NON_TER 513
CC FT 513 AA; 55704 MW; 943F31E3514CDEF3 CRC64;
SQ SEQUENCE
Alignment Scores:
Pred. No.: 2.53e-10 Length: 513
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.65% Indels: 0
Gaps: 0
DB: 1
US-09-873-224-147 (1-345) x POLG_HCVJ2 (1-513)
QY 51 CGGCCACAGAGCGTTAAGTCCAGCGCGGTGAGGAGTTAC 104
DB 18 ArgProGlnAspValIysPheProGlyGlyGlnIleValIglValTyr 35
RESULT 9
POLG_HCVH4 STANDARD: PRT: 520 AA.
ID POLG_HCVH4
AC 001404;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1)] (Fragment).
OS Hepatitis C virus (isolate HCV-476) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

CC Hepacivirus.
CC NCBI_taxid=31643;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=93019030; PubMed=1383400;
CC RX Abe K., Inchauste G., Fujisawa K.;
CC "Genomic characterization and mutation rate of hepatitis C virus
CC isolated from a patient who contracted hepatitis during an epidemic
CC of non-A, non-B hepatitis in Japan."
CC J. Gen. Virol. 73:2725-2729(1992).
CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MNNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10688; BAA01530.1; -
CC InterPro: IPR002531; HCV_NSI.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC Pfam: PF01539; HCV_core; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01560; HCV_NSI; 1.
CC ProDom: PD186062; HCV_NSI; 1.
CC Polyprotein: Glycoprotein; Coat protein; Envelope protein;
CC Transmembrane; Nonstructural protein.
CC INIT_MET 1
CC 1
CC CHAIN 1 115
CC CHAIN 116 191
CC CHAIN 192 383
CC CHAIN 384 >520
CC TRANSMEM 347 369
CC CARBOHYD 196 196
CC CARBOHYD 209 209
CC CARBOHYD 234 234
CC CARBOHYD 305 305
CC CARBOHYD 418 418
CC CARBOHYD 424 424
CC CARBOHYD 431 431
CC CARBOHYD 449 449
CC CARBOHYD 520 520
CC NON_TER 520
CC FT 520 AA; 56499 MW; AA135246CF20D525 CRC64;
SQ SEQUENCE
Alignment Scores:
Pred. No.: 2.53e-10 Length: 520
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.65% Indels: 0
Gaps: 0
DB: 1
US-09-873-224-147 (1-345) x POLG_HCVH4 (1-520)
QY 51 CGGCCACAGAGCGTTAAGTCCAGCGCGGTGAGGAGTTAC 104
DB 18 ArgProGlnAspValIysPheProGlyGlyGlnIleValIglValTyr 35
RESULT 10
POLG_HCVHK STANDARD: PRT: 520 AA.
ID POLG_HCVHK
AC 001403;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)

DE 16-OCT-2001 (Rel. 40, last annotation update)
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1)] (Fragment).
 OS Hepatitis C virus (isolate HCV-KE) (HCV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=31644;
 RN
 RN SEQUENCE FROM N.A.
 RA MEDLINE=93019030; PubMed=1383400;
 RA Abe K., Inchauspe G., Fujisawa K.;
 RT "Genomic characterization and mutation rate of hepatitis C virus
 RT isolated from a patient who contracted hepatitis during an epidemic
 RT of non-A, non-B hepatitis in Japan.";
 RT J. Gen. Virol. 73:2725-2729(1992).
 CC
 CC -i- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D10687; BAA01529.1; -
 DR PIR: J01925; J01925.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR Pfam: PF01539; HCV_env.1.
 DR Pfam: PF01542; HCV_core.1.
 DR Pfam: PF01543; HCV_capsid.1.
 DR Pfam: PF01560; HCV_NSI.1.
 DR ProDom: PD186062; HCV_NSI.1.
 DR PolyDome: Glycoprotein; Coat protein; Envelope protein;
 DR Transmembrane; Nonstructural protein.
 KW INIT_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 384 >520
 FT TRANSMEM 347 369
 FT TRANSMEM 196 196
 FT CARBOHYD 209 209
 FT CARBOHYD 233 233
 FT CARBOHYD 234 234
 FT CARBOHYD 305 305
 FT CARBOHYD 418 418
 FT CARBOHYD 424 424
 FT CARBOHYD 431 431
 FT CARBOHYD 449 449
 FT NON_TER 520 520
 SQ SEQUENCE 520 AA; 56476 MW; 1D2BD0A6FEF27349B CRC64;
 Alignment Scores:
 Pred. No.: 2,53e-10 Length: 520
 Score: 18.00 Matches: 18
 Percent Similarity: 100.00% Conservat: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 15.65% Indels: 0
 DB: 1 Gaps: 0
 US-09-873-224-147 (1-345) x POLG_HCVHK (1-520)
 OY 51 CGCCACAGACGTTAACTCCACAGCGCGGTGATGCTGGTGAATTAC 104
 ||||||||||||||||||||||||||||||||||||||||||||||||

DB 18 ArgProGlnAspValLysPheProGlyGlyGlnLeuValGlyGlyValTyr 35
 RESULT 11
 ID POLG_HCVBK STANDARD: PRT: 3010 AA.
 AC P26663;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Genome polyprotein [contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate BK) (HCV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CC NCBI_TaxID=11105;
 RN
 RN SEQUENCE FROM N.A.
 RA MEDLINE=9110698; PubMed=1847440;
 RA Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,
 RA Onishi E., Andoh T., Yoshida I., Okayama H.;
 RT "Structure and organization of the hepatitis C virus genome isolated
 RT from human carriers";
 RT J. Virol. 65:1105-1113(1991).
 RN
 RN [2]
 RP SEQUENCE OF 1487-1500.
 RX MEDLINE=96235224; PubMed=8647104;
 RX Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;
 RT "Non-structural protein 3 of hepatitis C virus inhibits
 RT phosphorylation mediated by cAMP-dependent protein kinase";
 RT Eur. J. Biochem. 237:611-618(1996).
 RN
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.
 RX MEDLINE=97015088; PubMed=8861916;
 RX Love R.A., Parag H.E., Wickersham J.A., Hostonsky Z., Habuka N.,
 RA Moonaw E.W., Adachi T., Hostonsky Z.;
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a
 RT trypsin-like fold and a structural zinc binding site";
 RL Cell 87:331-342(1996).
 RN
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.
 RX MEDLINE=98227846; PubMed=9568891;
 RA Van Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steinkuehler C., Tomei U., de Francesco R., Kuo L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C
 RT virus: a 2.2-A resolution structure in a hexagonal crystal form";
 RL Protein Sci. 7:837-847(1998).
 CC
 CC -i- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC
 CC -i- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC
 CC -i- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC [RNA] (N).
 CC
 CC -i- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC
 CC -i- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY 59.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL	M58335	AAA72945.1	-	
PIR	A38465	GNMVC		
PDB	1A10	25-MAR-98		
PDB	1JX3	14-JAN-98		
PDB	1NS3	08-APR-98		
DR	MEROPS	529.001	-	
DR	MEROPS	U39.001	-	
DR	InterPro	IPR001410	DEAD.	
DR	InterPro	IPR002531	HCV_NS1.	
DR	InterPro	IPR002518	HCV_NS2.	
DR	InterPro	IPR004109	HCV_NS3	
DR	InterPro	IPR00745	HCV_NS4a.	
DR	InterPro	IPR001490	HCV_NS4b.	
DR	InterPro	IPR002868	HCV_NS5a.	
DR	InterPro	IPR002166	HCV_RdRp.	
DR	InterPro	IPR002522	HCV_capsid.	
DR	InterPro	IPR002521	HCV_core.	
DR	InterPro	IPR002519	HCV_env.	
DR	Pfam	PF00998	HCV_RdRp	1.
DR	Pfam	PF01001	HCV_NS4b	1.
DR	Pfam	PF01006	HCV_NS4a	1.
DR	Pfam	PF01506	HCV_NS5a	1.
DR	Pfam	PF01538	HCV_NS2	1.
DR	Pfam	PF01539	HCV_env	1.
DR	Pfam	PF01542	HCV_core	1.
DR	Pfam	PF01543	HCV_capsid	1.
DR	Pfam	PF01560	HCV_NS1	1.
DR	Pfam	PF02907	HCV_NS3	1.
DR	ProDom	PD186062	HCV_NS1	1.
KW	Polyprotein	Glycoprotein	Transferase	RNA-directed RNA polymerase
KW	Core protein	Coat protein	Envelope protein	Helicase
KW	Transmembrane	Nonstructural protein	Hydrolase	Serine protease
KM	3D-structure			
FT	INT_MET	1	1	
FT	CHAIN	1	115	REMOVED FROM CAPSID PROTEIN C BY THE
FT	CHAIN	116	191	CELLULAR AMINOPEPTIDASE.
FT	CHAIN	192	383	CAPSID PROTEIN C (POTENTIAL).
FT	CHAIN	384	729	MATRIX PROTEIN C (POTENTIAL).
FT	CHAIN	730	1006	MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT	CHAIN	1007	1615	NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT	CHAIN	1616	1862	NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT	CHAIN	1863	2013	PROTEASE/HELICASE NS3 (POTENTIAL).
FT	CHAIN	2014	3010	NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT	CHAIN	3010	3010	NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT	CHAIN	347	369	RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT	TRANSMEM	1083	1083	POTENTIAL.
FT	ACT_SITE	1107	1107	CHARGE RELAY SYSTEM.
FT	ACT_SITE	1165	1165	CHARGE RELAY SYSTEM.
FT	ACT_SITE	1230	1237	CHARGE RELAY SYSTEM.
FT	NP_BIND	1316	1319	ATP (POTENTIAL).
FT	SITE	1316	1319	DECH BOX.
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	234	234	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	305	305	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	430	430	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	448	448	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	540	540	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2041	2041	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2077	2077	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2240	2240	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2529	2529	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2788	2788	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3010	3010	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	3010	327189	AA; 327189 MW; F8422D5ECCFDE9DC CRC64;

```

Pred. No.: 2, 07e-10 Length: 3010
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.65% Indels: 0
DB: 1 Gaps: 0

US-09-873-224-147 (1-345) x POLG_HCVBK (1-3010)
OY 51 CGGCCACGACGACGTTAAGTCCCGCGCGCGCGTCAAGTCGTGGAGTTTAC 104
D6 18 ArgProGlnAspValLysPheProGlyGlyGlyGlnIleValGlyGlyValTyr 35

RESULT 12
POLG_HCVJA
ID POLG_HCVJA STANDARD: PRT: 3010 AA.
AC P26662;
DR 01-AUG-1992 (Rel. 23, Created)
DR 01-AUG-1992 (Rel. 23, Last sequence update)
DR 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polypeptide [contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP25); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisvirio)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus
OX NCBI_TaxID=11116;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijioka M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugiyama T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RT [Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990)].
RN (2)
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=91192160; PubMed=1849488;
RA Kato N., Hijioka M., Nakagawa M., Ootsuyama Y., Muraio K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-328(1991).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC {RNA}(N).
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: D90208; BAAL4233.1; -.
DR PIR: A39253; GNMVCL.
DR HSSP: P26663; LJXP.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.

```


RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean C.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Slimmons M., Squares R., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymprez B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Beger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas R., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RA "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).

CC -! SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC LEPA SUBFAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: Z98598; CAB1233.1; -
 DR HSSP: P13551; IPNM.
 DR InterPro: IPR004161; EFTU_D2.
 DR InterPro: IPR000795; EF_GTPbind.
 DR InterPro: IPR005225; Small_GTP.
 DR Pfam: PF00009; GTP_EFTU; 1.
 DR Pfam: PF03144; GTP_EFTU_D2; 1.
 DR TIGRfams: TIGR00231; small_gtp; 1.
 DR PROSITE: PS00301; EFATOR_GTP; 1.
 KW Hypothetical protein; GTP-binding.
 FT NP_BIND 65 72 GTP (POTENTIAL).
 FT NP_BIND 130 134 GTP (POTENTIAL).
 FT NP_BIND 184 187 GTP (POTENTIAL).
 SO SEQUENCE 646 AA; 72683 MW; F9FA9498D384503E CRC64;

Alignment Scores:

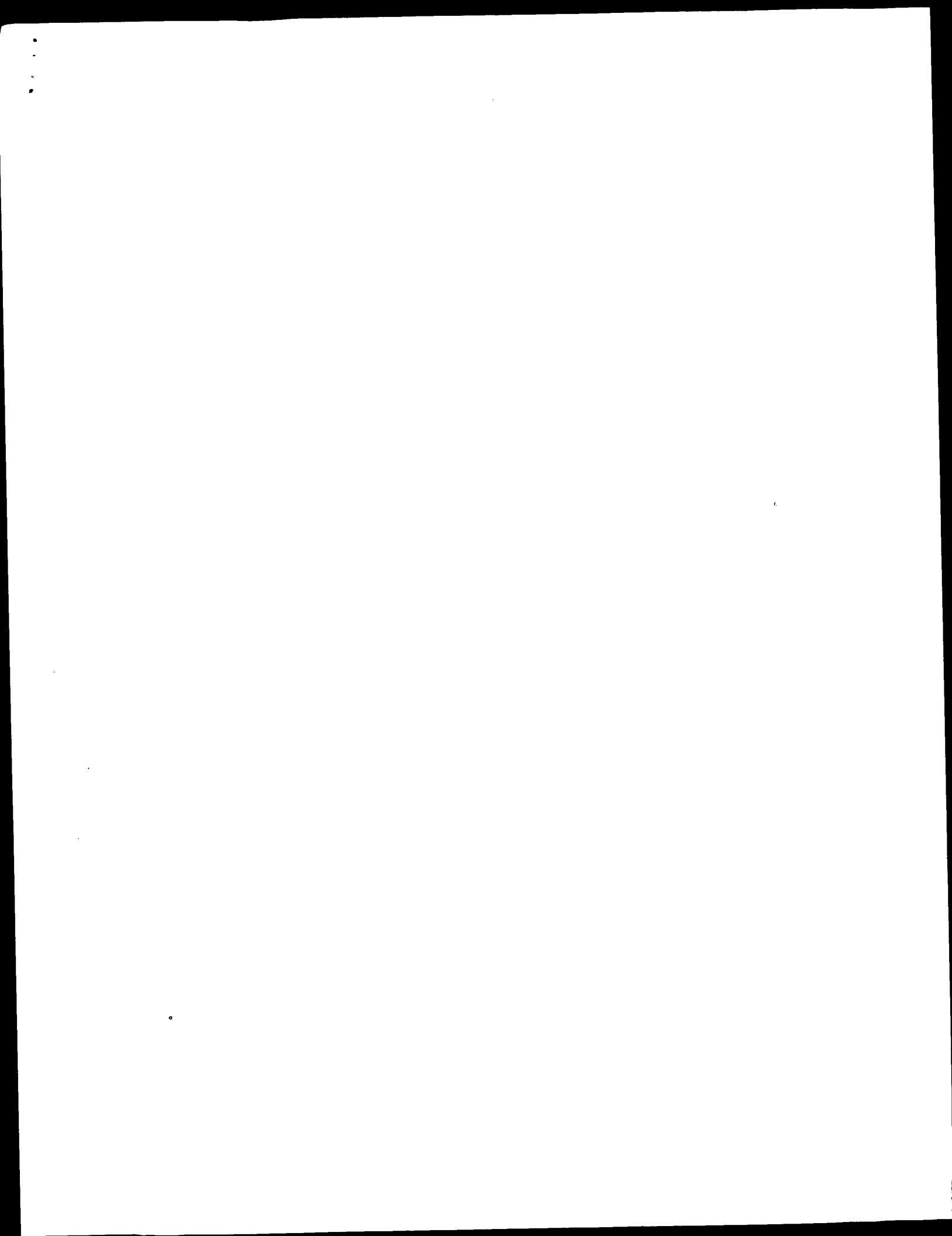
Pred. No.:	4.92	Length:	646
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	7.02%	Indels:	0
DB:	1	Gaps:	0

US-09-873-224-147 (1-345) x YE14_SCHPO (1-646)

OY 202 GCCCTGGGATGGGTGGCGCCTA 179

DB 385 AtAlaenglymetelYtrparglau 392

Search completed: February 19, 2003, 02:25:00
 Job time : 18 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

[illegible]

```
Title: US-09-873-224-147
Perfect score: 115
Sequence: 1 ATGAGCACACTTCTTAACC.....AAATGACCCCGGCGCAGGA 345
```

Scoring table:	ULL60
Xgapop	60.0 , Xgapext 60.0
Ygapop	60.0 , Ygapext 60.0
Fgapop	6.0 , Fgapext 7.0
Delop	6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343044

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

```

Command line parameters:
-MODEL=frame-m2p.model -DEV=xlp
-Q=/cgn2.1/USPTO/spool/US09873224/funalt_14022003_092759_28160/app_query.fasta_1.519
-BD=SPRREBL_21 -QMT=fastan -SUFFIX=0.1n2p.rspt -MINMATCH=0.1 -LOOPC=0
-LOOPEXT=0 -UNITs=bits -START=1 -END=1 -MATRIX=cligo -TRANS=humana.0.ccl
-LIST=45 -DOCALLIGN=200 -HSP SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09873224_e6gn_1_1_86_e7una_14022003_092759_28160 -MCPU=6 -LOCAL
-NO_XLPT -NO_MMAP -LARGESOURCE -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRAUDS=1 -XGAPOP=60 -XGAPEXT=60 -FEAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

```

```

SPRMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_potent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacterioph.*
17: sp_archaeop.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query			
No.	Score	Match	Length	ID
1	91	79.1	108	12 068494
				068494 hepatitis C

2	72	62.6	119	12	Q81294	Q81294	hepatitis c
3	72	62.6	415	12	Q68794	Q68794	hepatitis c
4	72	62.6	415	12	Q68799	Q68799	hepatitis c
5	72	62.6	415	12	Q68802	Q68802	hepatitis c
6	72	62.6	415	12	Q68808	Q68808	hepatitis c
7	72	62.6	3019	12	Q68801	Q68801	hepatitis c
8	65	56.5	415	12	Q68806	Q68806	hepatitis c
9	44	38.3	114	12	Q80Mj6	Q80mj6	hepatitis c
10	44	38.3	114	12	Q80Mj4	Q80mj4	hepatitis c
11	44	38.3	114	12	Q80Mj3	Q80mj3	hepatitis c
12	44	38.3	114	12	Q80Mj2	Q80mj2	hepatitis c
13	44	38.3	115	12	Q68898	Q68898	hepatitis c
14	44	38.3	120	12	Q80RM6	Q80rm6	hepatitis c
15	44	38.3	120	12	Q80RM5	Q80rm5	hepatitis c
16	44	38.3	120	12	Q80RM2	Q80rm2	hepatitis c
17	44	38.3	120	12	Q80RM1	Q80rm1	hepatitis c
18	44	38.3	120	12	Q80RM0	Q80rm0	hepatitis c
19	44	38.3	122	12	Q80Rj5	Q80rj5	hepatitis c
20	44	38.3	122	12	Q80Mj1	Q80mj1	hepatitis c
21	44	38.3	122	12	Q80Mj7	Q80mj7	hepatitis c
22	44	38.3	122	12	Q80Mj6	Q80mj6	hepatitis c
23	44	38.3	122	12	Q80Mj5	Q80mj5	hepatitis c
24	44	38.3	150	12	Q68861	Q68861	hepatitis c
25	44	38.3	150	12	Q68867	Q68867	hepatitis c
26	44	38.3	150	12	Q68863	Q68863	hepatitis c
27	44	38.3	150	12	Q68865	Q68865	hepatitis c
28	44	38.3	188	12	Q09740	Q09740	hepatitis c
29	44	38.3	191	12	Q68108	Q68108	hepatitis c
30	44	38.3	191	12	Q68109	Q68109	hepatitis c
31	44	38.3	191	12	Q68114	Q68114	hepatitis c
32	44	38.3	191	12	Q68125	Q68125	hepatitis c
33	44	38.3	191	12	Q68127	Q68127	hepatitis c
34	44	38.3	191	12	Q68152	Q68152	hepatitis c
35	44	38.3	191	12	Q68153	Q68153	hepatitis c
36	44	38.3	191	12	Q68155	Q68155	hepatitis c
37	44	38.3	191	12	Q68156	Q68156	hepatitis c
38	44	38.3	191	12	Q68157	Q68157	hepatitis c
39	44	38.3	191	12	Q68158	Q68158	hepatitis c
40	44	38.3	191	12	Q68162	Q68162	hepatitis c
41	44	38.3	195	12	Q91Tf7	Q91tf7	hepatitis c
42	44	38.3	195	12	Q91Tf6	Q91tf6	hepatitis c
43	44	38.3	319	12	Q81478	Q81478	hepatitis c
44	44	38.3	319	12	Q81263	Q81263	hepatitis c
45	44	38.3	415	12	Q81548	Q81548	hepatitis c

ALIGNMENTS

RESULT 1	068494	PRELIMINARY:	PRT:	108 AA.
ID	068494			
AC	068494;			
DT	01-NOV-1996 (Tremblurel. 01, Created)			
DT	01-NOV-1996 (Tremblurel. 01, Last sequence update)			
DT	01-JUN-2001 (Tremblurel. 17, Last annotation update)			
DE	Core region (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae			
NCBI_TaxID	11103;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97201609; PubMed=9049395;			
RA	van Doorn L.J., Kleter B.G.E.M., Stuyver L., Maetens G.,			
RA	Brouwer J.T., Schalm S.W., Helfink R.A., Quint W.G.V.,			
RT	"sequence analysis of Hepatitis C virus genotypes 1 to 5 reveals			
RT	multiple novel subtypes in the Benelux countries.";			
U	J. Gen. Virol. 76:1871-1876(1995).			
DR	EMBL: X7863; CA55412.1; -			
DR	InterPro: IPR002522; HCV_capsid.			
DR	Pfam: PF01543; HCV_capsid; 1.			
FT	NON_TER			
SQ	SEQUENCE	108 AA:	12170 MW:	953395BBIAD0A749E CRC64;

Alignment Scores:

Pred. No.:	3.12e-85	Length:	108
Score:	91.00	Matches:	91
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	79.13%	Indels:	0
DB:	12	Gaps:	0

US-09-873-224-147 (1-345) x Q68494 (1-108)

OY 51 CGGCCACAGACGTTAAGTTCAGGCGGCGTCAAGTCGTGAGCTTACGTCTA 110
 DB 18 ArgProGlnaPVallyspheProGlyGlnIleValGlyValTyrValIleu 37
 OY 111 CCACGACAGGGCCCCAGTTGGGTGGCGAGTCCGAGACTCCGAGCGTCA 170
 DB 38 ProArgArgGlyProGlnLeuGlyValArgAlaValArgLysThrSerGlnArgSerGln 57
 OY 171 CCGCAGTAGAGCGCCCAACCCATCCGAGGCGCGCGGAAACGAGGGGAGTCTGGCT 230
 DB 58 ProArgSerArgArgGlnProIleProArgAlaArgArgThrGlnArgSerTyrPala 77
 OY 231 CAGCCCGGATCCCTTGGCCCTATATGAGGATAGAGGCTGGGCGTGGAGGGTGC 290
 DB 78 GlnProGlyTyrProIleProIleProIleProIleProIleProIleProIleProIle 97
 OY 291 CCGTCCCGCGGCGCTCTGCGCGCGTGGTGGGCG 323
 DB 98 LeuSerProArgGlySerArgProSerTrpPly 108

RESULT 2

O81294 PRELIMINARY: PRT: 119 AA.
 ID 081294:
 AC 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Polyprotein (fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FD-3/93;
 RX MEDLINE=95189942; PubMed=7883898;
 RA Hotta H., Handajani R., Ingelund M., Soemarto W., Doi H.,
 Miyajima H., Homma M.;
 RT "Subtype analysis of hepatitis C virus in Indonesia on the basis of
 NS5b region sequences."
 RL J. Clin. Microbiol. 32:3049-3051(1994).
 DR EMBL: D30047; BAA06283.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 FT NON_TER 119
 FT SEQUENCE 119 AA; 13620 MW; 471715D6F84E58C8 CRC64;

Alignment Scores:

Pred. No.:	1.45e-65	Length:	119
Score:	72.00 <td>Matches: <td>72</td> </td>	Matches: <td>72</td>	72
Percent Similarity:	100.00% <td>Conservative: <td>0</td> </td>	Conservative: <td>0</td>	0
Best Local Similarity:	100.00% <td>Mismatches: <td>0</td> </td>	Mismatches: <td>0</td>	0
Query Match:	62.61% <td>Indels: <td>0</td> </td>	Indels: <td>0</td>	0
DB:	12	Gaps: <td>0</td>	0

US-09-873-224-147 (1-345) x Q81294 (1-119)

OY 129 TTGGGTGTGGTGCAGTGCAGACGCTCCGACGCTCGCAACTTCAGTAGCGCCAA 188
 DB 44 LeuGlyValArgAlaValArgLysThrSerGlnArgSerGlnProArgSerArgGln 63
 OY 189 CCCATCCCCAGGGCGCCGACGAGGCGAGGTCTCTGAGCTTCACCGCGGCTCTG 248

DB 64 ProIleProArgAlaArgArgThrGlnArgSerTrpPalaGlnProGlyTyrProTrp 83
 OY 249 CCCATATATGGAATGAGGCTCGCGGTGGCGAGGTGGCTCTCCCGCGGCTCT 308
 DB 84 ProLeuTyrGlyLysnGlnGlyCysGlyTrpAlaGlyTyrPleuLeuSerProArgGlySer 103
 OY 309 CCGCCGTCGTGGGGCCCAATGACCCCGCGGAG 344
 DB 104 ArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 3

O68794 PRELIMINARY: PRT: 415 AA.
 ID 068794:
 AC 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Genome polyprotein (Contains: envelope glycoprotein E2 (GP68) (GP70)
 DE (NS1) (fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JR030;
 RX MEDLINE=96226020; PubMed=8627233;
 RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F.,
 Lesmana L.A., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Jakarta, Indonesia classifiable into
 novel genotypes in the second (2e and 2f), tenth (10a) and eleventh
 (11a) genetic groups."
 RT J. Gen. Virol. 77:293-301(1996).
 CC -1- SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
 DR EMBL: D49747; BAA08581.1;
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; Transmembrane.
 FT NON_TER 415
 FT SEQUENCE 415 AA; 44797 MW; B9747B725A2B6238 CRC64;

Alignment Scores:

Pred. No.:	1.29e-65	Length:	415
Score:	72.00 <td>Matches: <td>72</td> </td>	Matches: <td>72</td>	72
Percent Similarity:	100.00% <td>Conservative: <td>0</td> </td>	Conservative: <td>0</td>	0
Best Local Similarity:	100.00% <td>Mismatches: <td>0</td> </td>	Mismatches: <td>0</td>	0
Query Match:	62.61% <td>Indels: <td>0</td> </td>	Indels: <td>0</td>	0
DB:	12	Gaps: <td>0</td>	0

US-09-873-224-147 (1-345) x Q68794 (1-415)

OY 129 TTGGGTGTGGTGCAGTGCAGACGCTCCGACGCTCGCAACTTCAGTAGCGCCAA 188
 DB 44 LeuGlyValArgAlaValArgLysThrSerGlnArgSerGlnProArgSerArgGln 63
 OY 189 CCCATCCCCAGGGCGCCGACGAGGCGAGGTCTCTGAGCTTCACCGCGGCTCTG 248
 DB 64 ProIleProArgAlaArgArgThrGlnArgSerTrpPalaGlnProGlyTyrProTrp 83
 OY 249 CCCATATATGGAATGAGGCTCGCGGTGGCGAGGTGGCTCTCCCGCGGCTCT 308
 DB 84 ProLeuTyrGlyLysnGlnGlyCysGlyTrpAlaGlyTyrPleuLeuSerProArgGlySer 103
 OY 309 CCGCCGTCGTGGGGCCCAATGACCCCGCGGAG 344

Db 104 ArgProSeTTPGlyProAsnAspProArgArg 115

RESULT 4

ID 068799 PRELIMINARY: PRT: 415 AA.

AC 068799:

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70) (NS1)] (Fragment).

DE Hepatitis C virus.

OS Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;

OC Hepacivirus.

OC NCBI_TaxID=11103;

RP SEQUENCE FROM N.A.

RP STRAIN-JK049;

RX MEDLINE=96226020; PubMed=8627233;

RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F.,

RA Lesmana L.A., Miyakawa Y., Mayumi M.,

RT "Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups."

RT J. Gen. Virol. 77:293-301(1996).

CC -i SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.

DR EMBL: D49749; BAA08583.1; -

DR InterPro: IPR002522; HCV_core.

DR InterPro: IPR002521; HCV_core.

DR InterPro: IPR002519; HCV_core.

DR InterPro: IPR002531; HCV_NSI.

DR Pfam: PF01543; HCV_core; 1.

DR Pfam: PF01542; HCV_core; 1.

DR Pfam: PF01539; HCV_core; 1.

DR Pfam: PF01560; HCV_NSI; 1.

DR ProDom: PD186062; HCV_NSI; 1.

DR Cost protein: Envelope protein: Glycoprotein: Nonstructural protein:

KW Polyprotein: Transmembrane.

FT NON_TER 415

SO SEQUENCE 415 AA; 44825 MM; 58C1C6B866E09F3 CRC64;

Alignment Scores:

Pred. No.:	1.29e-65	Length:	415
Score:	72.00	Matches:	72
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	62.61%	Indels:	0
DB:	12	Gaps:	0

US-09-873-224-147 (1-345) x Q68799 (1-415)

QY 129 TTGGGTGTCGTCAGTGCAGACTTCGAGCGGTCCGACCTGCGAGTGGCCCA 188

Db 44 LenglyValAlrghAlaValAlrghlsthserGluArgserGlnProArgSerArgArgGln 63

QY 189 CCATCCCCGAGGCGGCGGCGGAGGCGGAGGCTCTGGGCTCAGCCCGGCTGG 248

Db 64 ProleuTyrGlyAsnGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 83

QY 249 CCCCTATATGGAATGAGGCTGCGGCTGGGCGAGGCTGCTCTCCCGCGGCGCT 308

Db 84 ProleuTyrGlyAsnGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 103

QY 309 CGCCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 344

Db 104 ArgProSeTTPGlyProAsnAspProArgArg 115

RESULT 5

ID 068802 PRELIMINARY: PRT: 415 AA.

AC 068802:

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70) (NS1)] (Fragment).

OS Hepatitis C virus.

OS Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;

OC Hepacivirus.

OC NCBI_TaxID=11103;

RP SEQUENCE FROM N.A.

RP STRAIN-JK055;

RX MEDLINE=96226020; PubMed=8627233;

RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F.,

RA Lesmana L.A., Miyakawa Y., Mayumi M.,

RT "Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups."

RT J. Gen. Virol. 77:293-301(1996).

CC -i SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.

DR EMBL: D49750; BAA08584.1; -

DR InterPro: IPR002522; HCV_core.

DR InterPro: IPR002521; HCV_core.

DR InterPro: IPR002519; HCV_core.

DR InterPro: IPR002531; HCV_NSI.

DR Pfam: PF01543; HCV_core; 1.

DR Pfam: PF01542; HCV_core; 1.

DR Pfam: PF01539; HCV_core; 1.

DR Pfam: PF01560; HCV_NSI; 1.

DR Cost protein: Envelope protein: Glycoprotein: Nonstructural protein:

KW Polyprotein: Transmembrane.

FT NON_TER 415

SO SEQUENCE 415 AA; 44937 MM; D9008E5CBB58FF6B CRC64;

Alignment Scores:

Pred. No.:	1.29e-65	Length:	415
Score:	72.00	Matches:	72
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	62.61%	Indels:	0
DB:	12	Gaps:	0

US-09-873-224-147 (1-345) x Q68802 (1-415)

QY 129 TTGGGTGTCGTCAGTGCAGACTTCGAGCGGTCCGACCTGCGAGTGGCCCA 188

Db 44 LenglyValAlrghAlaValAlrghlsthserGluArgserGlnProArgSerArgArgGln 63

QY 189 CCATCCCCGAGGCGGCGGCGGAGGCGGAGGCTCTGGGCTCAGCCCGGCTGG 248

Db 64 ProleuTyrGlyAsnGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 83

QY 249 CCCCTATATGGAATGAGGCTGCGGCTGGGCGAGGCTGCTCTCCCGCGGCGCT 308

Db 84 ProleuTyrGlyAsnGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 103

QY 309 CGCCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 344

Db 104 ArgProSeTTPGlyProAsnAspProArgArg 115

RESULT 6

ID 068808 PRELIMINARY: PRT: 415 AA.

AC 068808:

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Genome polyprotein [Contains: envelope glycoprotein E2 (GP68) (GP70) (NS1)] (Fragment).

DE Hepatitis C virus.

OS Viruses; ssRNA positive-strand viruses, no DNA stage: Flaviviridae;

OC Hepacivirus.

OC NCBI_TaxID=11103;

RP SEQUENCE FROM N.A.

DR EMBL: D63821; BAA00980.1; -
DR HSSP: P27958; 1HEI.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR001917; NHRtransf_2.
DR Pfam: PF01543; HCV_capsid.1.
DR Pfam: PF01542; HCV_core.1.
DR Pfam: PF01539; HCV_env.1.
DR Pfam: PF01560; HCV_NS1.1.
DR Pfam: PF01538; HCV_NS2.1.
DR Pfam: PF02907; HCV_NS3.1.
DR Pfam: PF01006; HCV_NS4a.1.
DR Pfam: PF01001; HCV_NS4b.1.
DR Pfam: PF01506; HCV_NS5a.1.
DR Pfam: PF00998; HCV_RdRp.1.
DR Pfam: PF0186062; HCV_NS1.1.
DR Pfam: PF0186062; HCV_NS1.1.
DR PROSITE: PS00599; AA_TRANSFERR_CLASS_2; UNKNOWN.1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;
KW Transmembrane.
SQ SEQUENCE 3019 AA: 328210 MW: AF7A6774BC6D95FA CRC64;

Alignment Scores:			
Pred. No.:	1,06e-65	Length:	3019
Score:	72.00	Matches:	72
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	62.61%	Indels:	0
DB:	12	Gaps:	0
US-09-873-224-147 (1-345) x Q68801 (1-3019)			
QY	129 TTGGGTGTGCTGCAGTGGCGCAAGACTTCCAGCGGGTGCACACTCCGACGTAGAGCGCCAA	188	
Db	44 LKGLYVALARGALAAVALARGYSHRSERGLNARGSERGLNPROARGSERFARGGLN	63	
QY	189 CCCATCCCAGGGCGCGCGAACCAGGAGGAGTCCGTGAGCTCAGCCCGGATACCTTGG	248	
Db	64 PRLLEPRQARGALARGARGHRLNGLNARGSERTRPRLAAGLNPROGLTYTPROTIP	83	
QY	249 CCCCATATAGGAATAGAGAGGCTGCGGAGGAGGAGGTGCTGTCCCGCGGCTCT	308	
Db	84 PROLENTYGLYASNGLNLGYSELTYRPLAAGLYTPLEULAEUSERPROARGLYSER	103	
QY	309 CGCCCCGTCGTGGGGCCCAATGACCCCGCGCGCAG	344	
Db	104 ARGPROSETRTPGLYPROKSNAPROARGARG	115	
RESULT 8			
ID	Q68806		
AC	Q68806;		
	PRELIMINARY;	PRT,	415 AA.
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	Genome polypeptide [Contains: envelope glycoprotein E2 (GP68) (GP70)		
DE	(NS1)] (Fragment).		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;		
OC	Hepadnavirus.		
OX	NCBI_Taxid=11103;		
RN	[1]		

RP SEQUENCE FROM N.A.
 RC STRAIN=JK070;
 RX MEDLINE=96226020; PubMed=8627233;
 RA Tokita H., Okamoto H., Iizuka H., Kishimoto J., Tsuda F.,
 Iesmana L.A., Miyakawa Y., Mayumi M.;
 RT "Hepatitis C virus variants from Jakarta, Indonesia classifiable into
 RT novel genotypes in the second (2e and 2f), tenth (10a) and eleventh
 RT (11a) genetic groups";
 RT J. Gen. Virol. 77:293-301(1996).
 CC -1 SIMILARITY: TO HEPATITIS C VIRUS ENVELOPE GLYCOPROTEIN E1.
 DR EMBL: D49752; BAA08586.1; -;
 DR Interpro: IPR002522; HCV_capsid.
 DR Interpro: IPR002521; HCV_core.
 DR Interpro: IPR002519; HCV_env.
 DR Interpro: IPR002531; HCV_NSI.
 DR Pfam: PF01543; HCV_core; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polypeptide; Transmembrane.
 FT NON_TER 1 114
 SQ SEQUENCE 415 AA; 44979 MW; 3368538D667DEA39 CRC64;

Alignment Scores:
 Pred. No.: 2.28e-58 Length: 415
 Score: 65.00 Matches: 65
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 56.52% Indels: 0
 DB: Gaps: 0

US-09-873-224-147 (1-345) x O68806 (1-415)

OY 150 AAGACTCCGAGCGGTGCAACCTCGCAGTAGCGCCAACTCCCGAGCGCGCCGA 209
 |||||||
 DB 51 LysTrSerGluArgSerGlnProArgSerArgArgGlnProLeuProArgAlaArgArg 70
 OY 210 ACCGAGGCGAGTCTGCGCTCAGCCCGGTTACCTTGGCCCTATATGGGAATGAGGCG 269
 |||||||
 DB 71 ThrGluArgSerTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGluGly 90
 OY 270 TCGCGGTGGCAGGCGTCTCTCCCGCGCGGCTCTCGCCGTCGTGGGCGCCCAAT 329
 |||||||
 DB 91 CysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsn 110
 OY 330 GACCCCGCGCGCAGG 344
 |||||||
 DB 111 AspProArgArgArg 115

RESULT 9
 O80MJ6 PRELIMINARY; PRT; 114 AA.

ID O80MJ6
 AC 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Polypeptide (Fragment).
 DE Hepatitis C virus.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepatitis C virus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=110253A;
 RA Theamboonlers A., Bedi K., Scott N., Sriponthong M., Poovorawan Y.,
 RT "Hepatitis C virus genotyping by RFLP and direct sequencing in
 RT Thailand";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY089772; AAM09932.1; -;
 FT NON_TER 1 114
 SQ SEQUENCE 114 AA; 12938 MW; 16BF76D73C4BA811 CRC64;

Alignment Scores:
 Pred. No.: 1.43e-36 Length: 114
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: Gaps: 0

US-09-873-224-147 (1-345) x O80MJ6 (1-114)

OY 213 GAGGCGAGTCTCTGCGCTCAGCCCGGTTACCTTGGCCCTATATGGGAATGAGGCTGC 272
 |||||||
 DB 68 GluArgSerTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGluGlyCys 87
 OY 273 GGTGGCGAGGCGTCTCTCTCCCGCGCGGCTCTCGCCGTCGTGGGCGCCCAATGAC 332
 |||||||
 DB 88 GlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 107
 OY 333 CCGCGCGCAGG 344
 |||||||
 DB 108 ProArgArgArg 111

RESULT 11
 O80MJ3 PRELIMINARY; PRT; 114 AA.

ID O80MJ3
 AC 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Polypeptide (Fragment).
 DE Hepatitis C virus.
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepatitis C virus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=087203A;
 RA Theamboonlers A., Bedi K., Scott N., Sriponthong M., Poovorawan Y.,
 RT "Hepatitis C virus genotyping by RFLP and direct sequencing in
 RT Thailand";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY089774; AAM09934.1; -;
 FT NON_TER 1 114
 SQ SEQUENCE 114 AA; 12962 MW; 82EB67C69345338 CRC64;

Alignment Scores:
 Pred. No.: 1.43e-36 Length: 114
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: Gaps: 0

DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DE Polypeptide (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 NCBI_TaxID=11103;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=086293A;
 RA Theamboonlers A., Bedi K., Scottaj N., Sripontong M., Poovorawan Y.;
 RT "Hepatitis C virus genotyping by RFLP and direct sequencing in
 Thailand";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY089775; AAM09935.1; -;
 FT NON_TER 1 1
 SQ SEQUENCE 114 AA; 12971 MW; F3067192B507571E CRC64;

Alignment Scores:
 Pred. No.: 1.43e-36 Length: 114
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x Q8QMJ3 (1-114)

QY 213 GAGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 272
 DB 68 GAGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 272
 QY 273 GGGTGGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 332
 DB 88 GAGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 332
 QY 333 CCCCAGGCGAGG 344
 DB 108 Proargargarg 111

RESULT 12

ID Q8QMJ2 PRELIMINARY; PRT: 114 AA.
 AC Q8QMJ2;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DE Polypeptide (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 NCBI_TaxID=11103;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=152013A;
 RA Theamboonlers A., Bedi K., Scottaj N., Sripontong M., Poovorawan Y.;
 RT "Hepatitis C virus genotyping by RFLP and direct sequencing in
 Thailand";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY089776; AAM09936.1; -;
 FT NON_TER 1 1
 SQ SEQUENCE 114 AA; 12660 MW; 4CE7D826DABDC4EE CRC64;

Alignment Scores:
 Pred. No.: 1.43e-36 Length: 114
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x Q8QMJ2 (1-114)

QY 213 GAGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 272
 DB 68 GAGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 272
 QY 273 GGGTGGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 332
 DB 88 GAGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 332
 QY 333 CCCCAGGCGAGG 344
 DB 108 Proargargarg 111

RESULT 13

ID Q68898 PRELIMINARY; PRT: 115 AA.
 AC Q68898;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DE (genotype 4, N3) core region RNA (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 NCBI_TaxID=11103;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97201609; PubMed=9049395;
 RA van Doorn L.J., Kleter B.G.E.M., Stuyver L., Maertens G.,
 RA Brouwer J.T., Schalm S.W., Heijlink R.A., Quint W.G.V.;
 RT "Sequence analysis of Hepatitis C virus genotypes 1 to 5 reveals
 multiple novel subtypes in the Benelux countries";
 RL J. Gen. Virol. 76:1871-1876(1995).
 RN (2)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96048319; PubMed=8551256;
 RA Kleter B.G.E.M., van Doorn L.J., Stuyver L., Maertens G., Brouwer J.T.,
 RA Schalm S.W., Heijlink R.A., Quint W.G.V.;
 RT "Rapid genotyping of Hepatitis C virus RNA-isolates obtained from
 patients residing in Western Europe";
 RL J. Med. Virol. 47:35-42(1995).
 DR EMBL; Z29467; CAA82605.1; -;
 DR InterPro: IP002522; HCV_capsid.
 DR Pfam: PF01543; HCV_capsid; 1.
 FT NON_TER 115
 SQ SEQUENCE 115 AA; 13084 MW; CF03A686002D7CA8 CRC64;

Alignment Scores:
 Pred. No.: 1.43e-36 Length: 115
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x Q68898 (1-115)

QY 213 GAGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 272
 DB 72 GAGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 272
 QY 273 GGGTGGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 332
 DB 92 GGGTGGGCGAGGCTCTGGCTGACCCGGGTACCTGGCCCTATATGGAATGAGGCTGC 332
 QY 333 CCCCAGGCGAGG 344
 DB 112 Proargargarg 115

RESULT 14

Q8QRM6

ID 080RM6 PRELIMINARY; PRT; 120 AA.
 AC 080RM6;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE Polypeptide (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PS;
 RA Theamboonlers A., Bedi K., Scottaj N., Sriponthong M., Poovorawan Y.,
 RT "Hepatitis C virus genotyping by RFLP and direct sequencing in
 Thailand.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF482727; AAL82639.1; -.
 FT NON_TER 1 1
 FT CHAIN 7 >120 CORE PROTEIN.
 FT NON_TER 120 120
 SQ SEQUENCE 120 AA; 13562 MW; 73A8EF4925267A77 CRC64;

Alignment Scores:
 Pred. No.: 1.42e-36 Length: 120
 Score: 44.00 Matches: 44
 Percent Similarity: 100.00% Conservat: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x 080RM6 (1-120)

QY 213 GAGGCGAGTCTGGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 272
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 66 GlUGlYArgSerTrpAlaGlnProGlyTyrProTrrProLeuTrrGlyAsnGluGlyCys 85
 QY 273 GGGTGGGCGAGGCTGCTCTGTCGCCCGGCGGCTCTCGCCCGTGGGGCCCAATGAC 332
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 86 G1YTrpAlaG1YTrpLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 105
 QY 333 CCGCGGCGCAGG 344
 ||||||||||||
 Db 106 ProArgArgArg 109

RESULT 15

ID 080RM5 PRELIMINARY; PRT; 120 AA.
 AC 080RM5;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE Polypeptide (Fragment).
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NS;
 RA Theamboonlers A., Bedi K., Scottaj N., Sriponthong M., Poovorawan Y.,
 RT "Hepatitis C virus genotyping by RFLP and direct sequencing in
 Thailand.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF482728; AAL82640.1; -.
 FT NON_TER 1 1
 FT CHAIN 7 >120 CORE PROTEIN.
 FT NON_TER 120 120
 SQ SEQUENCE 120 AA; 13657 MW; D3ACF3264C199BDA CRC64;

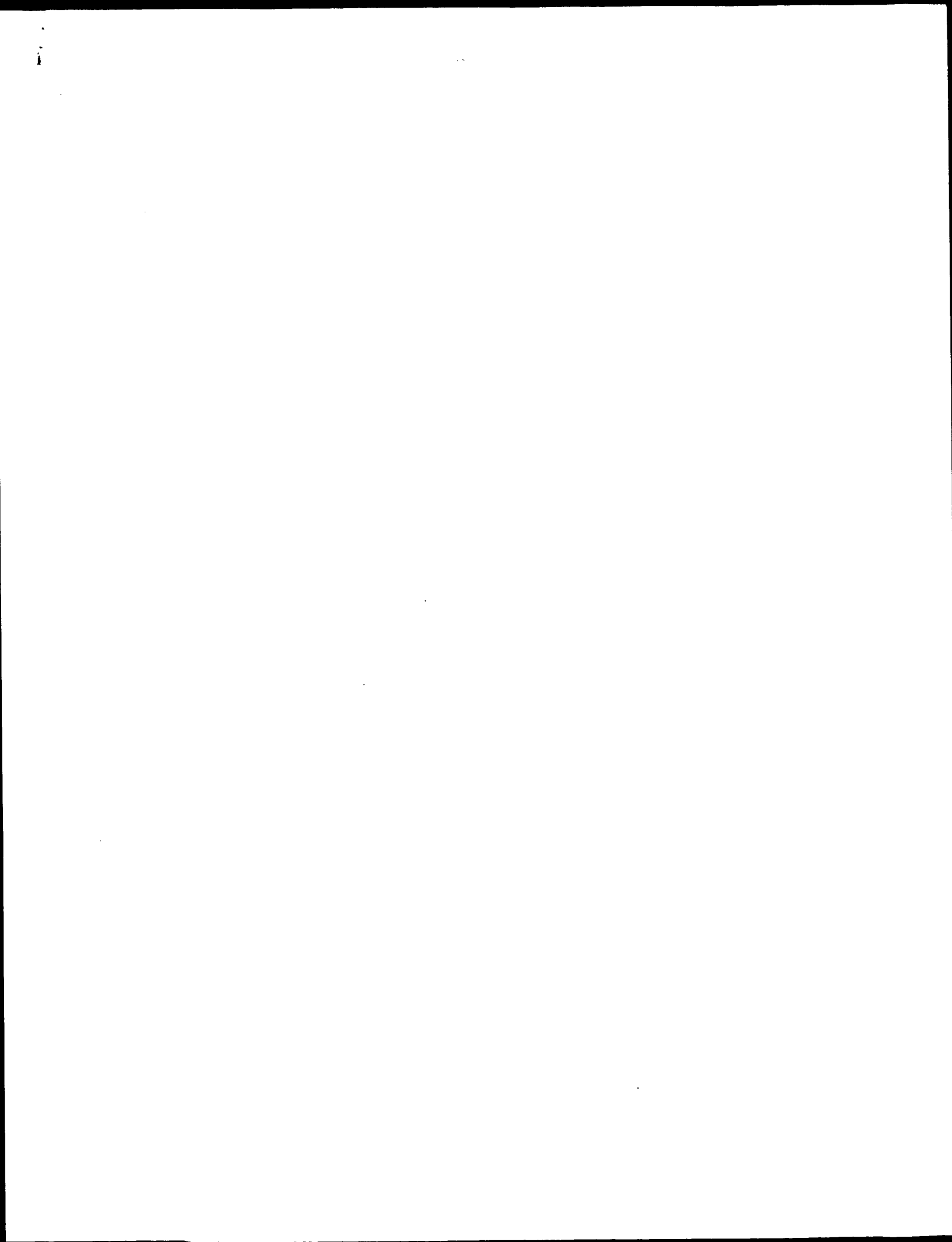
Alignment Scores:
 Pred. No.: 1.42e-36 Length: 120
 Score: 44.00 Matches: 44

Percent Similarity: 100.00% Conservat: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.26% Indels: 0
 DB: 12 Gaps: 0

US-09-873-224-147 (1-345) x 080RM5 (1-120)

QY 213 GAGGCGAGTCTGGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 272
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 66 GlUGlYArgSerTrpAlaGlnProGlyTyrProTrrProLeuTrrGlyAsnGluGlyCys 85
 QY 273 GGGTGGGCGAGGCTGCTCTGTCGCCCGGCGGCTCTCGCCCGTGGGGCCCAATGAC 332
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 86 G1YTrpAlaG1YTrpLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 105
 QY 333 CCGCGGCGCAGG 344
 ||||||||||||
 Db 106 ProArgArgArg 109

Search completed: February 19, 2003, 02:26:07
 Job time : 32 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 19, 2003, 02:23:06 ; Search time 14.5 Seconds

(without alignments)
1400.125 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 115
Sequence: 1 ATATGACACACTTCTTAACC.....AATATGACCCGCGCGACGA 345

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+.n2p.model -DEV=xip
-Q=/cgn2_1/USPRO.spool/US0873224/runat_14022003.092800.28231/app.query.fasta_1.519
-DB=Issued.Patents_AA -OFMT=fasta -SUFFIX=olin2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-OUTFMT=pio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US0873224.qcgn1_1_21.qrunat_14022003.092800.28231 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

Issued.Patents_AA:*
1: /cgn2_6/prodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/iaa/5A.COMB.pep:*
4: /cgn2_6/prodata/1/iaa/5B.COMB.pep:*
5: /cgn2_6/prodata/1/iaa/5A.COMB.pep:*
6: /cgn2_6/prodata/1/iaa/5B.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	85.2	115	4	US-08-836-075A-50
2	44	38.3	124	1	US-08-244-116B-15
3	44	38.3	191	2	US-08-290-665A-187
4	44	38.3	191	2	US-08-290-665A-188
5	44	38.3	191	2	US-08-290-665A-189
6	44	38.3	191	2	US-08-290-665A-190
7	44	38.3	191	2	US-08-290-665A-191
8	44	38.3	191	2	US-08-290-665A-192
9	44	38.3	191	2	US-08-290-665A-193
10	44	38.3	191	2	US-08-290-665A-195
11	44	38.3	191	2	US-08-290-665A-196
12	44	38.3	191	2	US-08-290-665A-197

13	44	38.3	191	5	PCT-US95-10398-187	Sequence 187, App
14	44	38.3	191	5	PCT-US95-10398-188	Sequence 188, App
15	44	38.3	191	5	PCT-US95-10398-189	Sequence 189, App
16	44	38.3	191	5	PCT-US95-10398-190	Sequence 190, App
17	44	38.3	191	5	PCT-US95-10398-191	Sequence 191, App
18	44	38.3	191	5	PCT-US95-10398-192	Sequence 192, App
19	44	38.3	191	5	PCT-US95-10398-193	Sequence 193, App
20	44	38.3	191	5	PCT-US95-10398-195	Sequence 195, App
21	44	38.3	191	5	PCT-US95-10398-196	Sequence 196, App
22	44	38.3	191	5	PCT-US95-10398-197	Sequence 197, App
23	37	32.2	191	5	US-08-290-665A-194	Sequence 194, App
24	37	32.2	191	5	PCT-US95-10398-194	Sequence 194, App
25	34	29.6	42	4	US-08-380-160-10	Sequence 10, Appl
26	34	29.6	46	1	US-08-262-037-27	Sequence 27, Appl
27	34	29.6	56	1	US-08-262-037-28	Sequence 28, Appl
28	34	29.6	61	1	US-08-262-037-29	Sequence 29, Appl
29	34	29.6	89	1	US-07-681-703B-24	Sequence 24, Appl
30	34	29.6	89	2	US-08-407-410B-24	Sequence 24, Appl
31	34	29.6	89	2	US-08-485-500-24	Sequence 24, Appl
32	34	29.6	89	5	PCT-US91-02370-24	Sequence 24, Appl
33	34	29.6	119	1	US-07-681-703B-18	Sequence 18, Appl
34	34	29.6	119	2	US-08-407-410B-18	Sequence 18, Appl
35	34	29.6	119	5	US-08-485-500-18	Sequence 18, Appl
36	34	29.6	119	5	PCT-US91-02370-18	Sequence 18, Appl
37	34	29.6	120	4	US-08-850-328-2	Sequence 2, Appl
38	34	29.6	144	4	US-08-444-818-103	Sequence 103, App
39	34	29.6	150	1	US-07-681-703B-16	Sequence 16, Appl
40	34	29.6	150	2	US-08-407-410B-16	Sequence 16, Appl
41	34	29.6	150	2	US-08-485-500-16	Sequence 16, Appl
42	34	29.6	150	5	PCT-US91-02370-16	Sequence 16, Appl
43	34	29.6	169	4	US-08-444-818-93	Sequence 93, Appl
44	34	29.6	190	4	US-08-078-271B-1	Sequence 1, Appl
45	34	29.6	191	2	US-08-290-665A-155	Sequence 155, App

ALIGNMENTS

RESULT 1
US-08-836-075A-50
Sequence 50, Application US/08836075A
Patent No. 6180768
GENERAL INFORMATION:
APPLICANT: MAERTENS, GEERT
APPLICANT: STUYVER, LIEVEN
TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
TITLE OF INVENTION: AGENTS
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON
STATE: TEXAS
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:

NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-075A-50

Alignment Scores:
Pred. No.: 2.74e-83 Length: 115
Score: 98.00 Matches: 98
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 85.22% Indels: 0
DB: Gaps: 0

US-09-873-224-147 (1-345) x US-08-836-075A-50 (1-115)

QY 51 CGGCCAGAGAGCTTACGTCACAGCGCGGTCAGATCGTTGGTGAAGTTACGTGCTA 110
DB 18 ATGPTGlnaspVallyPheProclgylglnlleValgIlyVallyr-Valleu 37
QY 111 CCAGCCAGGGGCCCCCAGTGGGTGTGCGTCAGTGCACAAAGACTTCGACGCGTCCAA 170
DB 38 ProlArgArglyProglInleuglyAlarAlaValArglystInSerGlnArgSerGln 57
QY 171 CCTGCGAGAGCGCCCAACCCATCCCGAGGGGCGCGACGAGGAGGCGCGGCT 230
DB 58 ProlArgSerArgArgGlnProleProArgAlaArgAlaArgIngluGlyArgSerTrpAla 77
QY 231 CAGCCGCGGTACCTCTGGCCCTATATGGAATGAGAGGCTGCGGGTGGCGAGGGTGCCTC 290
DB 78 GlnProglYTrpProTrpProleuTrpGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeu 97
QY 291 CTTGCCCCCGGCGCTCTGCGCCGTCGTCGCGGCGCCCAATGACCCCGGCGCAGG 344
DB 98 LeuSerProArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 2
US-08-244-116B-15
Sequence 15, Application US/08244116B
Patent No. 5763159

GENERAL INFORMATION:
APPLICANT: Simmonds, Peter
APPLICANT: Chan, Shiu-Wan
APPLICANT: Yap, Peng L.
TITLE OF INVENTION: Hepatitis-C Virus Testing
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
STREET: 1211 East Morehead Street
CITY: Charlotte
STATE: No. 5763159th Carolina
COUNTRY: United States
ZIP: 28234
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/244,116B
FILING DATE: 15-JUL-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/GB92/02143
FILING DATE: 20-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Sibley, Kenneth D.
REGISTRATION NUMBER: 31,665

REFERENCE/DOCKET NUMBER: 1749-125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 704-377-1561
TELEFAX: 704-334-2014
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: yes
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Hepatitis-C virus
US-08-244-116B-15

Alignment Scores:
Pred. No.: 4.22e-33 Length: 124
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: Gaps: 0

US-09-873-224-147 (1-345) x US-08-244-116B-15 (1-124)

QY 213 GAGGCGAGGTCTCTGCGTACGCCCGGCTACCTTGGCCCTATATGGAATGAGGCGTGC 272
DB 68 GlnGlyArgSerTrpAlaGlnProglYTrpProTrpProleuTrpGlyAsnGluGlyCys 87
QY 273 GGGGCGAGGAGTGGCTCTGCTCCCGCGGCGCTTCGCGCCGTGCGGGCCCAATGAC 332
DB 88 GlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 107
QY 333 CCGCGCGCAGG 344
DB 108 ProlArgArgArg 111

RESULT 3
US-08-290-665A-187
Sequence 187, Application US/08290665A
Patent No. 5882852

GENERAL INFORMATION:
APPLICANT: BURKH, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: HK10
US-08-290-665A-187

Alignment Scores:
Pred. No.: 3,96e-33 Length: 191
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x US-08-290-665A-187 (1-191)

QY 213 GAGGCGAGGCTCTGGGCTGACCCGGTACCTTGCCCTATATGGAATGAGGCTGC 272
Db 72 GAGGCGAGGCTCTGGGCTGACCCGGTACCTTGCCCTATATGGAATGAGGCTGC 272

QY 273 GGGTGGCGAGGCTGCTGCTGCCCGCGGCTCTGCGCGCTGCGGCGCCCAATGAC 332
Db 92 GGTTPALAGLYTTPleuLeuSerProArGlySerArGProSerTrpGlyProAsnsp 111

QY 333 CCCCAGCGCAGG 344
Db 112 ProArGArGArG 115

RESULT 4
US-08-290-665A-188
Sequence 188, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 188:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S52
US-08-290-665A-188

Alignment Scores:
Pred. No.: 3,96e-33 Length: 191
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x US-08-290-665A-188 (1-191)

QY 213 GAGGCGAGGCTCTGGGCTGACCCGGTACCTTGCCCTATATGGAATGAGGCTGC 272
Db 72 GAGGCGAGGCTCTGGGCTGACCCGGTACCTTGCCCTATATGGAATGAGGCTGC 272

QY 273 GGGTGGCGAGGCTGCTGCTGCCCGCGGCTCTGCGCGCTGCGGCGCCCAATGAC 332
Db 92 GGTTPALAGLYTTPleuLeuSerProArGlySerArGProSerTrpGlyProAsnsp 111

QY 333 CCCCAGCGCAGG 344
Db 112 ProArGArGArG 115

RESULT 5
US-08-290-665A-189
Sequence 189, Application US/08290665A
Patent No. 5882852
GENERAL INFORMATION:
APPLICANT: BURK, J., MILLER, R.H. AND
APPLICANT: PURCELL, R.H.
TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 189:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids

TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: S2
US-08-290-665A-189

Alignment Scores:
Pred. No.: 3,96e-33 Length: 191
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x US-08-290-665A-189 (1-191)

QY 213 GAGGCGAGTCTGCTGCTACGCCGGGACCTTGCCCTATATGGAATGAGGCTGC 272
DB 72 GUGLYARSGERTPALAGLNPGLTYRPTOTRPROLEUTYRGLYASNGLGLCY 91

QY 273 GGGTGGGCGAGGTGCTGCTGCCCGCGGCTCTCGCCCTCGTGGGCCCCAATGAC 332
DB 92 GLTYRPAAGLYTRPLEULEUSERPROARGLYSERARGPROSETRTPGLYPROASNAP 111

QY 333 CCCCAGGCGCAGG 344
DB 112 PROARGARGARG 115

RESULT 6
US-08-290-665A-190
Sequence 190, Application US/08290665A
Patent No. 5882852

GENERAL INFORMATION:
APPLICANT: BURKH, J., MILLER, R.H. AND
TITLE OF INVENTION: NOCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 190:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown

ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: DK12
US-08-290-665A-190

Alignment Scores:
Pred. No.: 3,96e-33 Length: 191
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x US-08-290-665A-190 (1-191)

QY 213 GAGGCGAGTCTGCTGCTACGCCGGGACCTTGCCCTATATGGAATGAGGCTGC 272
DB 72 GUGLYARSGERTPALAGLNPGLTYRPTOTRPROLEUTYRGLYASNGLGLCY 91

QY 273 GGGTGGGCGAGGTGCTGCTGCCCGCGGCTCTCGCCCTCGTGGGCCCCAATGAC 332
DB 92 GLTYRPAAGLYTRPLEULEUSERPROARGLYSERARGPROSETRTPGLYPROASNAP 111

QY 333 CCCCAGGCGCAGG 344
DB 112 PROARGARGARG 115

RESULT 7
US-08-290-665A-191
Sequence 191, Application US/08290665A
Patent No. 5882852

GENERAL INFORMATION:
APPLICANT: BURKH, J., MILLER, R.H. AND
TITLE OF INVENTION: NOCLEOTIDE AND DEDUCED
TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
NUMBER OF SEQUENCES: 263
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 191:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: 24

US-08-290-665A-191

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
3,96e-33	44.00	100.00%	100.00%	38.26%	191	44	0	0	0	0

DB: 2

US-09-873-224-147 (1-345) x US-08-290-665A-191 (1-191)

OY 213 GAGGCGAGTCTGAGCTACCCGGGTACCTTGCCCTATATGGGAATGAGGCTGC 272

Db 72 GUGlyArgSerTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyAsnGlyCys 91

OY 273 GGGTGGGCGAGGTCCTGCTGTCGCCGCGCTCTGCCGCTGTGAGGCCCAATGAC 332

Db 92 GLyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111

OY 333 CCCCAGCGCAGG 344

Db 112 ProArgArgArg 115

RESULT 8

US-08-290-665A-192

; Sequence 192, Application US/08290665A

; Patent No. 5882852

GENERAL INFORMATION:

APPLICANT: BORK, J., MILLER, R.H. AND

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,665A

FILING DATE: 15-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 751-6849

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 192:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: Z8

US-08-290-665A-192

Alignment Scores:

Pred. No.:

Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
3,96e-33	44.00	100.00%	100.00%	38.26%	191	44	0	0	0

DB: 2

US-09-873-224-147 (1-345) x US-08-290-665A-192 (1-191)

OY 213 GAGGCGAGTCTGAGCTACCCGGGTACCTTGCCCTATATGGGAATGAGGCTGC 272

Db 72 GUGlyArgSerTrpAlaGlnProGlyTrpProTrpProLeuTrpGlyAsnGlyCys 91

OY 273 GGGTGGGCGAGGTCCTGCTGTCGCCGCGCTCTGCCGCTGTGAGGCCCAATGAC 332

Db 92 GLyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp 111

OY 333 CCCCAGCGCAGG 344

Db 112 ProArgArgArg 115

RESULT 9

US-08-290-665A-193

; Sequence 193, Application US/08290665A

; Patent No. 5882852

GENERAL INFORMATION:

APPLICANT: BORK, J., MILLER, R.H. AND

TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

NUMBER OF SEQUENCES: 263

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/290,665A

FILING DATE: 15-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: RICHARD W. BORK

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4116

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 751-6849

TELEFAX: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 193:

SEQUENCE CHARACTERISTICS:

LENGTH: 191 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

ORIGINAL SOURCE:

ORGANISM: homosapiens

INDIVIDUAL ISOLATE: Z1

US-08-290-665A-193

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Length:	Matches:	Conservative:
3,96e-33	44.00	100.00%	191	44	0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x US-08-290-665A-193 (1-191)

QY 213 GAGGCGAGTCTGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 272
|||||
Db 72 GAGGCGAGTCTGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 272
|||||

QY 273 GAGGCGAGTCTGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 332
|||||
Db 92 GAGGCGAGTCTGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 332
|||||

QY 333 CCCCCGCGCAG 344
|||||
Db 112 ProArGArGArG 115

RESULT 10

US-08-290-665A-195
; Sequence 195, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z6
US-08-290-665A-195

Alignment Scores:
Pred. No.: 3,96e-33 Length: 191
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x US-08-290-665A-195 (1-191)

QY 213 GAGGCGAGTCTGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 272
|||||
Db 72 GAGGCGAGTCTGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 272
|||||

QY 273 GAGGCGAGTCTGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 332
|||||
Db 92 GAGGCGAGTCTGGCTCAGCCCGGGTACCTTGCCCTATATGGAATGAGGCTGC 332
|||||

QY 333 CCCCCGCGCAG 344
|||||
Db 112 ProArGArGArG 115

RESULT 11

US-08-290-665A-196
; Sequence 196, Application US/08290665A
; Patent No. 5882852
; GENERAL INFORMATION:
; APPLICANT: BURKH, J., MILLER, R.H. AND
; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
; TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
; TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
; NUMBER OF SEQUENCES: 263
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,665A
FILING DATE: 15-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4116
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 196:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
ORIGINAL SOURCE:
ORGANISM: homosapiens
INDIVIDUAL ISOLATE: Z7
US-08-290-665A-196

Alignment Scores:
Pred. No.: 3,96e-33 Length: 191
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
DB: 2 Gaps: 0

US-09-873-224-147 (1-345) x US-08-290-665A-196 (1-191)

US-09-873-224-147 (1-345) x PCT-US95-10398-189 (1-191)

OY	213	GAAGGAGGTCCTCCGAGGCGGCTACCCCTATGAGGAATGAGGCTGC	272
Db	72	GlUGlyArgSerTrpAlaGlnProGlyTrpProTrpProLeuTrcLYAsnGluGlyCys	91
OY	273	GAGTGGGAGGAGGCTCCCTCCGAGGCGGCTCTCGCCCGTGGGCGCCAAATGAC	332
Db	92	GlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyProAsnAsp	111
OY	333	CCCCGGGCGCAGG	344
Db	112	ProArgArgArg	115

Search completed: February 19, 2003, 02:27:35
 Job time : 16.5 secs

GenCore version 5.1.4_P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: February 19, 2003, 02:24:25; Search time 11 Seconds
(without alignments)
1602.611 Million cell updates/sec

Title: US-09-873-224-147

Perfect score: 115
Sequence: 1 AATGACACACTTCTTAAC.....AATGACCCCGCGCAGCA 345

Scoring table:
OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 140259 seqs, 2554876 residues

Word size: 1

Total number of hits satisfying chosen parameters: 255142

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+np.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09873224/runat_14022003_092800_28261/app.query.fasta.1.519
-DB=Published.Applications_AA -QMT=fastan -SUFFIX=clin2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFM=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09873224.@CGN_1.1_12.@runat_14022003_092800_28261
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_XLPXY=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database: Published.Applications_AA.*

1: /cgn2.6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2.6/ptodata/1/pubpaa/PTI_NEW_PUB.pep.*
3: /cgn2.6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2.6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2.6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2.6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2.6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2.6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2.6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2.6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2.6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2.6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2.6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	85.2	115	9	US-09-851-138-50
2	98	85.2	115	9	US-09-899-046-148
3	44	38.3	166	9	US-09-899-046-164
4	36	31.3	166	9	US-09-899-046-194

5	34	29.6	113	10	US-09-921-397-78	Sequence 78, Appl
6	34	29.6	126	9	US-09-899-046-166	Sequence 166, App
7	34	29.6	182	9	US-10-104-966-2	Sequence 2, Appl
8	34	29.6	182	10	US-09-929-955-2	Sequence 2, Appl
9	34	29.6	2894	10	US-09-941-611-23	Sequence 23, Appl
10	34	29.6	3011	9	US-09-747-419-20	Sequence 20, Appl
11	34	29.6	3011	9	US-10-104-966-1	Sequence 1, Appl
12	34	29.6	3011	9	US-09-891-894-3	Sequence 3, Appl
13	34	29.6	3011	9	US-09-995-937-20	Sequence 20, Appl
14	34	29.6	3011	10	US-09-742-659-4	Sequence 4, Appl
15	34	29.6	3011	10	US-09-916-355-2	Sequence 2, Appl
16	34	29.6	3011	10	US-09-238-076-20	Sequence 20, Appl
17	34	29.6	3011	10	US-09-952-572-9	Sequence 9, Appl
18	34	29.6	3011	10	US-09-929-955-1	Sequence 1, Appl
19	34	29.6	3012	9	US-09-995-937-2	Sequence 2, Appl
20	34	29.6	3012	10	US-09-238-076-2	Sequence 2, Appl
21	30	26.1	319	9	US-09-851-138-44	Sequence 44, Appl
22	30	26.1	319	9	US-09-851-138-48	Sequence 48, Appl
23	28	24.3	109	9	US-09-851-138-6	Sequence 6, Appl
24	28	24.3	319	9	US-09-851-138-12	Sequence 12, Appl
25	28	24.3	319	9	US-09-851-138-42	Sequence 42, Appl
26	25	21.7	74	9	US-09-851-138-10	Sequence 10, Appl
27	25	21.7	96	9	US-09-899-046-192	Sequence 192, Appl
28	25	21.7	137	9	US-09-851-138-46	Sequence 46, Appl
29	25	21.7	138	9	US-09-851-138-60	Sequence 60, Appl
30	23	20.0	109	9	US-09-851-138-2	Sequence 2, Appl
31	23	20.0	117	9	US-09-851-138-28	Sequence 28, Appl
32	22	19.1	319	9	US-09-851-138-18	Sequence 18, Appl
33	19	16.5	19	9	US-09-899-046-268	Sequence 268, App
34	19	16.5	44	10	US-09-880-945-10	Sequence 10, Appl
35	19	16.5	158	9	US-09-851-138-66	Sequence 66, Appl
36	18	15.7	74	9	US-09-851-138-104	Sequence 104, App
37	18	15.7	77	10	US-09-921-397-3	Sequence 3, Appl
38	18	15.7	91	10	US-09-758-308-1	Sequence 1, Appl
39	18	15.7	97	10	US-09-756-875-8	Sequence 8, Appl
40	18	15.7	103	10	US-09-921-397-77	Sequence 77, Appl
41	18	15.7	108	9	US-09-851-138-14	Sequence 14, Appl
42	18	15.7	139	10	US-09-736-959A-17	Sequence 17, App
43	18	15.7	166	9	US-09-899-046-152	Sequence 152, App
44	18	15.7	169	9	US-09-899-046-42	Sequence 42, App
45	18	15.7	169	9	US-09-899-046-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-09-851-138-50
; Sequence 50, Application US/09851138
; Publication No. US20020183508A1

GENERAL INFORMATION:

APPLICANT: MAERTENS, GEERT
STUYVER, LIEVEN

TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESSES:

ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. BOX 4433
CITY: HOUSTON

STATE: TEXAS

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/851,138
FILING DATE: 09-May-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,075

FILING DATE: <Unknown>
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-09-851-138-50

Alignment Scores:
Pred. No.: 7.58e-83 Length: 115
Score: 98.00 Matches: 98
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 85.22% Indels: 0
Gaps: 0
DB: 9

US-09-873-224-147 (1-345) x US-09-851-138-50 (1-115)

QY 51 CGGCCACAGAGGTTAAGTTCACAGCGCGGTCCAGATCCTTGGTGAAGTTACGTGCTA 110
|||||
DB 18 ArgProGlnAspValLysPheProGlyGlyGlnIleValGlyValAlaGlyTyrValLeu 37
QY 111 CCACGAGAGGCGCCCAAGTTGGGTGTCGTCAGTCGCAAGACTTCCGAGCGGTGCA 170
|||||
DB 38 ProArgArgGlyProGlnLeuGlyValAlaValAlaGlyThrSerGlnArgSerGln 57
QY 171 CTTGCCAGTAGGCGCCCAACCATCCCGAGGCGCGCCGAAACGAGGAGTCCTGGGCT 230
|||||
DB 58 ProArgSerArgArgGlnProIleProArgAlaArgArgThrGlnGlyArgSerTrpAla 77
QY 231 CAGCCGGGTACCTTGCGCCCTATATGGAATGAGGCGGTGGGCGAGGGTGGCTC 290
|||||
DB 78 GlnProGlyTyrProIleProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeu 97
QY 291 CTGTCCCGCGCGGCTCTCGCCCGTCGTGGGGGCCCAATGACCCCGCGCGAG 344
|||||
DB 98 LeuSerProArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 2
US-09-899-046-148
Sequence 148, Application US/09899046
Publication No. US2003008274A1
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus
TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,046
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 148:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-09-899-046-148

Alignment Scores:
Pred. No.: 7.58e-83 Length: 115
Score: 98.00 Matches: 98
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 85.22% Indels: 0
Gaps: 0
DB: 9

US-09-873-224-147 (1-345) x US-09-899-046-148 (1-115)

QY 51 CGGCCACAGAGGTTAAGTTCACAGCGCGGTCCAGATCCTTGGTGAAGTTACGTGCTA 110
|||||
DB 18 ArgProGlnAspValLysPheProGlyGlyGlnIleValGlyValAlaGlyTyrValLeu 37
QY 111 CCACGAGAGGCGCCCAAGTTGGGTGTCGTCAGTCGCAAGACTTCCGAGCGGTGCA 170
|||||
DB 38 ProArgArgGlyProGlnLeuGlyValAlaValAlaGlyThrSerGlnArgSerGln 57
QY 171 CTTGCCAGTAGGCGCCCAACCATCCCGAGGCGCGCCGAAACGAGGAGTCCTGGGCT 230
|||||
DB 58 ProArgSerArgArgGlnProIleProArgAlaArgArgThrGlnGlyArgSerTrpAla 77
QY 231 CAGCCGGGTACCTTGCGCCCTATATGGAATGAGGCGGTGGGCGAGGGTGGCTC 290
|||||
DB 78 GlnProGlyTyrProIleProLeuTyrGlyAsnGlnGlyCysGlyTrpAlaGlyTrpLeu 97
QY 291 CTGTCCCGCGCGGCTCTCGCCCGTCGTGGGGGCCCAATGACCCCGCGCGAG 344
|||||
DB 98 LeuSerProArgGlySerArgProSerTrpGlyProAsnAspProArgArgArg 115

RESULT 3
US-09-899-046-164
Sequence 164, Application US/09899046
Publication No. US2003008274A1
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus
TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,046
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 164:
SEQUENCE CHARACTERISTICS:
LENGTH: 166 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-899-046-164

Alignment Scores:
Pred. No.: 6.25e-33 Length: 166
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.26% Indels: 0
Gaps: 0
DB: 9

US-09-873-224-147 (1-345) x US-09-899-046-164 (1-166)

QY 213 GAGGCGAGGTCTGTGGGCTCAGCCCGGTACCTTGCCCTATATGGAATGAGGCGCTGC 272
|||||


```

      LENGTH: 113
      TYPE: PRT
      ORGANISM: Hepatitis C virus
US-09-921-397-78

Alignment Scores:
Pred. No.: 1,16e-23 Length: 113
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.57% Indels: 0
DB: 10 Gaps: 0

US-09-873-224-147 (1-345) x US-09-921-397-78 (1-113)
QY 225 TGGCGTCAGCCCGGCTACCTTGCCCTATATGGGAATGAGCGCTCGCGGTGGCGACGG 284
      |||
Db 76 TrrpladlnrrhoglyrrProtrrrProleuTyrgLYAsnGLUGLYCysGLYrrPladGLY 95
QY 285 TGGCTCTGTCCCCGCGGCGCTCTGCGCCCGTGGTGGGGCCCA 326
      |||
Db 96 TrpleuLeuSerProArGlySerArqProSerTrpGLYPro 109

RESULT 6
US-09-899-046-166
; Sequence 166, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 166:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-09-899-046-166

Alignment Scores:
Pred. No.: 1,14e-23 Length: 126
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.57% Indels: 0
DB: 9 Gaps: 0

US-09-873-224-147 (1-345) x US-09-899-046-166 (1-126)
QY 201 GCGCGCGCAACGACGAGGCGAGGTCTGGGGCTACCGCGGTACCTTGCCCTATATGG 260
      |||
Db 68 AlArArqrrgrhrrLUGLYArSerTrrrPladGLInrrProGLYrrProTrrrProleuTyrgLY 87
QY 261 AATGAGGCTGCGGCTGGCGGCGACGGTGGCTCTGTCCCCGCG 302
      |||
Db 88 AsnGLUGLYCysGLYrrPladGLYrrPlauLeuSerProArq 101

RESULT 7
US-10-104-966-2
; Sequence 2, Application US/10104966

```

Patent No. US20020155124A1
GENERAL INFORMATION:
APPLICANT: Matti Sallberg
APPLICANT: Catharina Hultgren
TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: TRIPEP.23AUSCI
CURRENT APPLICATION NUMBER: US/10/104,966
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: 09/705,547
PRIOR FILING DATE: 2000-11-03
PRIOR APPLICATION NUMBER: 60/229,175
PRIOR FILING DATE: 2000-08-29
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Hepatitis C virus core protein sequence
US-10-104-966-2

Alignment Scores:
Pred. No.: 1.09e-23 Length: 182
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.57% Indels: 0
DB: 9 Gaps: 0

US-09-873-224-147 (1-345) x US-10-104-966-2 (1-182)

QY 225 TGGGCTCAGCCCGGCTACCCCTTGGCCCTATATGGAATAGAGGCTCGGGTGGCAGG 284
Db 76 TTPALAGINPROGLTYTPROTTPROLEUHYRGLYASNLUGLYCSGLYTPPALAGLY 95
QY 285 TGGCTCTGTCGCCGGGCTCTGCCCGCTGCGGCGCCCA 326
Db 96 TTPLEULSETPROARGLYSERARGPROSERTTPGLYPRO 109

RESULT 8
US-09-929-955-2
Sequence 2, Application US/09929955
Patent No. US20020136740A1
GENERAL INFORMATION:
APPLICANT: Matti Sallberg
APPLICANT: Catharina Hultgren
TITLE OF INVENTION: VACCINES CONTAINING RIBAVIRIN AND
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: TRIPEP.23AUS2
CURRENT APPLICATION NUMBER: US/09/929,955
CURRENT FILING DATE: 2001-08-15
PRIOR APPLICATION NUMBER: 09/705,547
PRIOR FILING DATE: 2000-11-03
PRIOR APPLICATION NUMBER: 60/229,175
PRIOR FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: 60/225,767
PRIOR FILING DATE: 2000-08-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Hepatitis C virus core protein sequence
US-09-929-955-2

Alignment Scores:
Pred. No.: 1.09e-23 Length: 182
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.57% Indels: 0
DB: 10 Gaps: 0

US-09-873-224-147 (1-345) x US-09-929-955-2 (1-182)

QY 225 TGGGCTCAGCCCGGCTACCCCTTGGCCCTATATGGAATAGAGGCTCGGGTGGCAGG 284
Db 76 TTPALAGINPROGLTYTPROTTPROLEUHYRGLYASNLUGLYCSGLYTPPALAGLY 95
QY 285 TGGCTCTGTCGCCGGGCTCTGCCCGCTGCGGCGCCCA 326
Db 96 TTPLEULSETPROARGLYSERARGPROSERTTPGLYPRO 109

RESULT 9
US-09-941-611-23
Sequence 23, Application US/09941611
Patent No. US20020106640A1
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
POLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSMUN, HUGO

TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.

STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VA

COUNTRY: USA
ZIP: 22201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611

FILING DATE: 30-Aug-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671

FILING DATE: 1995-02-21

APPLICATION NUMBER: WO PCT/EP91/02409

FILING DATE: 13-DEC-1991

APPLICATION NUMBER: EP 90124241.2

FILING DATE: 14-DEC-1990

ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B. J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 1487-5

TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 2894 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-941-611-23

Alignment Scores:
Pred. No.: 7.71e-24 Length: 2894
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0


```

; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/995,937
; FILING DATE: 28-May-1998
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,756
; FILING DATE: 04-May-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: HOLLAND, DONALD R.
; REGISTRATION NUMBER: 35,197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 314-727-5188
; TELEFAX: 314-727-6092
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; FRAGMENT TYPE: N-terminal
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-995-937-20

Alignment Scores:
Pred. No.: 7.67e-24 Length: 3011
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.57% Indels: 0
DB: Gaps: 0

US-09-873-224-147 (1-345) x US-09-995-937-20 (1-3011)
QY 225 TGGGCTCAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGCGGCTGGCGCAGG 284
DB 76 TTPALAGINProGlyTyrProTrrProLeuTyrGlyAsnGlyCysGlyTrrpalaGly 95
QY 285 TGGCTCTGTCCCGCGGCGCTCGCCCGTGGTGGGCGCCA 326
DB 96 TTPLeuLeuSerProArgGlySerArgProSerTrrpGlyPro 109

RESULT 14
US-09-742-659-4
; Sequence 4, Application US/09742659
; Patent No. US20010034019A1
; GENERAL INFORMATION:
; APPLICANT: Hong, Zhi
; APPLICANT: Butkiewicz, Nancy J.
; APPLICANT: Zhong, Weidong
; APPLICANT: Ingravallo, Paul
; APPLICANT: Wright-Minogue, Jacquelyn
; APPLICANT: Lau, Johnson Y.
; APPLICANT: Lemon, Stanley M.
; TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
; FILE REFERENCE: ID01116
; CURRENT APPLICATION NUMBER: US/09/742,659
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/171,469
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 16

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-742-659-4

Alignment Scores:
Pred. No.: 7.67e-24 Length: 3011
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.57% Indels: 0
DB: Gaps: 0

US-09-873-224-147 (1-345) x US-09-742-659-4 (1-3011)
QY 225 TGGGCTCAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGCGGCTGGCGCAGG 284
DB 76 TTPALAGINProGlyTyrProTrrProLeuTyrGlyAsnGlyCysGlyTrrpalaGly 95
QY 285 TGGCTCTGTCCCGCGGCGCTCGCCCGTGGTGGGCGCCA 326
DB 96 TTPLeuLeuSerProArgGlySerArgProSerTrrpGlyPro 109

RESULT 15
US-09-916-359-2
; Sequence 2, Application US/09916359
; Patent No. US20020034734A1
; GENERAL INFORMATION:
; APPLICANT: Veronique Barban
; TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
; FILE REFERENCE: PMCE97-03A
; CURRENT APPLICATION NUMBER: US/09/916,359
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 09/388,874
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: 97/02,887
; PRIOR FILING DATE: 1997-03-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Virus
US-09-916-359-2

Alignment Scores:
Pred. No.: 7.67e-24 Length: 3011
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.57% Indels: 0
DB: Gaps: 0

US-09-873-224-147 (1-345) x US-09-916-359-2 (1-3011)
QY 225 TGGGCTCAGCCCGGCTACCTTGCCCTATATGGAATGAGGCTGCGGCTGGCGCAGG 284
DB 76 TTPALAGINProGlyTyrProTrrProLeuTyrGlyAsnGlyCysGlyTrrpalaGly 95
QY 285 TGGCTCTGTCCCGCGGCGCTCGCCCGTGGTGGGCGCCA 326
DB 96 TTPLeuLeuSerProArgGlySerArgProSerTrrpGlyPro 109

Search completed: February 19, 2003, 02:28:13
Job time : 16 secs

```